ISSN 1864-6168

A DERIVATION OF THE MEAN ABSOLUTE DISTANCE IN ONE-DIMENSIONAL RANDOM WALK

Hižak J. [1], **Logožar R.** [1]
[1]**Polytechnic of Varaždin, Varaždin, Croatia**

***Abstract**. In this paper we argue on the use of the mean absolute deviation in 1D random walk as opposed to the commonly accepted standard deviation. It presents an in detail derivation of the closed-form formula for the 1D mean absolute distance, including the proof by induction. The limit for the infinite number of steps is included.*

***Key words:** random walk, Brownian movement, mean absolute deviation, mean (expected) value of the absolute deviation (distance).*

***Sažetak**. U ovom članku razmatramo uporabu srednjeg apsolutnog odstupanja za 1D nasumični hod, nasuprot opće prihvaćenoj standardnoj devijaciji. Članak daje detaljni izvod eksplicitne formule za srednju apsolutnu udaljenost za 1D nasumični hod, uključujući dokaz indukcijom i limes u graničnom slučaju kada je broj koraka beskonačan.*

***Ključne riječi:** nasumični hod, Brownovo gibanje, srednje apsolutno odstupanje, srednja (očekivana) vrijednost apsolutnog odstupanja (udaljenosti).*

1. INTRODUCTION

In today's science the notions of *the mean square deviation* or *variance*, and the associated *standard deviation* (SD) present the very foundation of the statistics used in all spheres of natural and technical sciences. Standard deviation is a widely accepted and almost unavoidable research tool of every experimental and theoretical scientist. The formula for standard deviation is one of the first that students acquire in the statistical courses.

In the well known problem of random walk, a common approach is to use the squares of the distances from the starting point and to calculate its mean value [1,2,3].This is equivalent to the concept of finding the variance of a certain probability distribution, and presents the classic result of the statistical physics.

However, the fact often not revealed, or simply neglected, is that the quadratic dispersion measures are not the only possible. The other, intuitively simpler and easier to understand, is *mean absolute deviation* (MAD). It is a simple average of the absolute deviations, or differences of the set elements from its mean value. Since it is so intuitive, the question arises why it is not used more often? And what would this measure give in the case of the random walk?

During the last century several authors, mostly from the field of statistics, argued that for the dispersion measure and, then also for the equivalent problem of random walk, the absolute values could be used as instead of the squared [4,5]. Speaking of the mean absolute deviation, the famous mathematician R.A. Fisher admits that "for some types of work it is more expeditious than the use of squares" [5]. However, it is surprising that for such a common statistical and physical topic, there is little or no feedback, and no comparative analysis of the two dispersion measures in textbooks and literature.

The aim of our work is to contribute to the clarification of this important topic. We hope to synthesize the arguments which will approve the use of quadratic measures from the deeper physical grounds. As a "first step" in the journey, in this paper we calculate the mean absolute distance of 1D random walk, and compare it to the classic results for variance $\sigma^2 = n$, and standard deviation $\sigma = \sqrt{n}$, after n number of steps.

2. RANDOM WALK

In general, the random walk is a mathematical model that describes any motion consisting of a number of random steps. This model can be applied to numerous phenomena in economics (fluctuating stock price), game theory (financial status of a gambler), or even biology (the search path of a foraging animal). The classical example of a 2-D random walk is the Brownian movement: the motion of a particle on a liquid surface, induced by collisions with the nearby molecules. After n collisions, of which each has transferred to the particle a momentum of random direction, and, in the most general case, of random amount, the trajectory will be a collection of n successive random steps. By placing the particle inside a medium, we get to the more general 3-D random walk.

2.1 Basic definitions of 1D random walk

In this paper we deal with the simplest case, the *one dimensional* (1D) *random walk*. An object, initially positioned at $x = \mu$, starts to stroll along the x-axis with two possible movement choices at any given point: i) to step left or ii) to step right. We define the position $x(n)$ of the object in the n-th step. In the simplest case the steps are of equal length d, and the probabilities of stepping left or right are equal.

According to this, after the first step the object's position $x(1)$ is given by:

$$x(1) = \begin{cases} \mu - d\,, & \text{1st step to the left}\,, \\ \mu + d\,, & \text{1st step to the right}\,. \end{cases}$$

After the second step the position is:

ISSN 1864-6168

$$x(2) = \begin{cases} \mu + 2d\,, & \text{1st step left, 2nd left,} \\ \mu\,, & \text{1st step left, 2nd right,} \\ \mu\,, & \text{1st step right, 2nd left,} \\ \mu + 2d\,, & \text{1st step right, 2nd right.} \end{cases}$$

Immediately we spot the dichotomy between the odd and even number of steps which will follow us throughout our calculations. In the same way we will analyze the distance $x(n)$ after arbitrary number n of steps. A possible random walk path is illustrated in Figure 1.

The *distance* $\Delta x(n)$ from the starting point μ in the ***n***-th step is:

$$\Delta x(n) = x(n) - \mu\,. \tag{1}$$

Statistically, this distance corresponds to the *deviation* from the mean value. So, the two terms are synonyms and we shall use both of them as appropriate.

After defining the deviation or distance, we are ready to pose the central statistical questions for our system: what is the mean or expected value of the object's position and what is the dispersion: given as standard deviation and the mean absolute deviation.

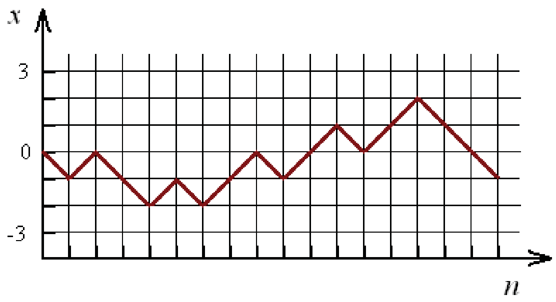

Figure 1. One possible path in the 1D random walk. The horizontal coordinate *n* is the number of steps taken; the vertical coordinate x is the distance from the starting position

2.2 Mean and expected Value

To make our discussion complete, we shall briefly address the simple facts on the object's position mean or expected values. Regarding the difference between ***mean*** and ***expected***, one must notice that in a strict, formal approach it is obviously important. The *mean value* is the result of the calculation over the empirically collected data,

$$\bar{x} = \sum_{k=1}^{n} f_{rel_k} x_k\,. \tag{2}$$

On the other hand, the *expected value* stands for the weighted average of all the possible values that a random variable can have, with each value being weighted by its assumed apriori probability [6]:

$$E(x) = \sum_{k=1}^{n} p_k x_k\,. \tag{3}$$

Very often the subtle difference between the two notions and the corresponding values are ignored. In the strict sense, we will use the apriori probabilities of the binomial distribution, and thus derive the expected value. However, by taking the physical reality as a starting point, the notion of apriori probabilities is just an idealization of what should be ultimately tested in an experiment. In other words, in the real world we always start from some relative frequencies and calculate the mean values, and the (apriori) probabilities can be interpreted as the relative frequencies of a certain outcome x_k with a large number m of trials. In this case the relative frequencies go into the probabilities, and the mean value into the expected value:

$$f_{rel\,k,m} \xrightarrow[m\to\infty]{} p_{k,m}\ ,\ \bar{x}_m \xrightarrow[m\to\infty]{} E(x)\,. \tag{4}$$

So, from the physical standpoint, there is no essential difference between the mean and expected value. We shall mostly use the term ***mean*** to depict both notions, even when calculating it from the apriori probabilities from a probability distribution. Furthermore, we shall use the brackets $\langle\ \ \rangle$ to present the averaging operator, so that in further text the following simplification is assumed: $\langle x \rangle = \bar{x} = E(x)$.

Thus for the *mean* (*expected*) *distance* in the 1D random walk we write simply:

$$\langle \Delta x(n) \rangle = \langle x(n) - \mu \rangle\ . \tag{5}$$

Since the averaging operator is linear, it turns out that the above mean distance is zero:

$$\langle \Delta x(n) \rangle = \langle x(n) \rangle - \langle \mu \rangle = \langle x(n) \rangle - \mu \equiv 0\,. \tag{6}$$

The result is "expected", indeed. Since there is no preferred direction of random walk, the probabilities to go left or right are equal.

In order to simplify the problem we shall consider the case in which the walk starts at point $\mu = 0$ and at each move takes a step $d = -1$ or $d = +1$. The simplification leads us to:

$$\Delta x(n) = x(n) - 0 = x(n),$$
$$\langle \Delta x(n) \rangle = \langle x(n) \rangle = 0\,. \tag{7}$$

2.3 Variance and standard deviation

We have just repeated the trivial result that the mean value of the linear deviation vanishes. The mean (expected) value of the deviation squared is generally nonzero, and presents the well known *variance* σ^2. It corresponds to the average distance squared $\langle x^2(n) \rangle$, so that we have:

$$\sigma^2 = \langle x^2(n) \rangle = \sum_{k=1}^{n} p_k (x_k - \mu)^2\ , \tag{8}$$

$$\mu = 0\ \Rightarrow\ \sigma^2 = \langle x^2(n) \rangle = \sum_k p_k x_k\,. \tag{9}$$

In the second equation, a shorthanded of the summing notation is introduced that will be used in the rest of the text: the summing will always go up to the index n.

Now we are ready to find out the mean square distance $\langle x^2(n) \rangle$ for the 1D random walk. We follow the elegant inductive derivation by Feynman [1]. The distance for the first step is $x(n) = x(1) = \pm 1$, leading to the unique squared value $x^2 = 1$. With the equal probability of taking + or − direction, the mean squared distance is:

$$\langle x^2(1) \rangle = \frac{1}{2}\cdot(-1)^2 + \frac{1}{2}\cdot 1^2 = \frac{1}{2} + \frac{1}{2} = 1\,.$$

The net distance after $(n-1)$ steps is $x(n-1)$, so that after the next, n-th step, there are two possible, equally probable outcomes: $x(n)=x(n-1)-1$ or $x(n)=x(n-1)+1$. For their squares we have the following:

$$x^2(n)=\begin{cases}x^2(n-1)-2\cdot x(n-1)+1\\ or\\ x^2(n-1)+2\cdot x(n-1)+1\end{cases} \tag{10}$$

Since both directions are equally probable, the average value is obtained in the same as before, leading to the cancelation of the mid terms and the average squared distance:

$$\langle x^2(n)\rangle=\frac{1}{2}(2\langle x^2(n-1)\rangle+2)$$
$$=\langle x^2(n-1)\rangle+1\,. \tag{11}$$

We have already shown that $\langle x^2(1)\rangle=1$. From there it follows that:

$$\langle x^2(2)\rangle=\langle x^2(1)\rangle+1=2\,,\ldots\,,\langle x^2(n-1)\rangle=n-1\,,$$
$$\langle x^2(n)\rangle=n\,, \tag{12}$$

which is the well known text-book result.

Besides the mean squared deviation, we may be interested in the corresponding *linear indicator*, the standard deviation σ or the *root-mean-square distance* value:

$$\sigma=x_{RMS}=\sqrt{\langle x^2(n)\rangle}\,. \tag{13}$$

The RMS value is of utmost importance in physics and technical sciences, having many interpretations in both, discrete and continuous domain. In our case of 1D random walk the standard deviation, or RMS value, of the distance is:

$$x_{RMS}=\sqrt{n}\,. \tag{14}$$

3. RANDOM WALK VIA BINOMIAL PROBABILITY DISTRIBUTION

Here we shall expose a more general approach to the results (6) and (12) that will also serve us as a preparation for the calculation of the mean absolute distance in the next section.

3.1 Random walk in Pascal's triangle

We introduce the Pascal's triangle as an excellent way of visualizing the random walk (Figure 2). Here we can follow the development of the random walk as the number of steps rises. Each node in the triangle represents a possible ending point of the random walk path. At height (dept) n, there are in total $n+1$ nodes. Each node can be uniquely denoted as the ordered pair (n,k), with $n=0,1,2,\ldots$, and $k=0,1,\ldots,n$, starting from left to right. To each node its corresponding binomial coefficient $\binom{n}{k}$ is attributed. It shows the number of possible paths leading to that node, and is written in the Pascal's triangle as usual.

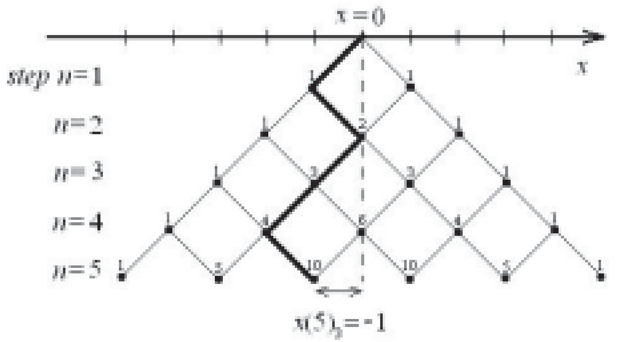

Figure 2 Random walk in the Pascal's triangle. The bold line represents the first five steps of the path shown in Figure 1. The number of possible paths ending in a node is the same as the corresponding binomial coefficient.

We can immediately note that for odd n the middle nodes are missing, which is the consequence of the fact that the walk can end in the zero distance point only after an even number steps.

For example, if the number of steps is $n=4$, there are in all 16 possible paths, of which 1 ends at $x=-4$, 4 paths end at $x=-2$, and 6 at zero point $x=0$. The exactly symmetrical situation is for the positive distances. The corresponding binomial coefficients $\binom{n}{k}$ have values $k=0,1$ for the negative distances $(x=-4,-2)$, the value $k=2$ for the zero distance $(x=0)$, and values $k=3,4$ for the positive distances $(x=2,4)$.

The probability that a random path ends in a node can be described as a node probability. To get it for the node k at the height n we must divide the corresponding binomial coefficient $\binom{n}{k}$ with the total number of paths for that height: $\sum_{k=1}^{n}\binom{n}{k}=2^n$.

In our example of $n=4$ steps, the central node corresponding to $x=0$ has the index $k=2$, and the probability that the random walk ends here is:

$$p_{4,2}=\frac{1}{2^4}\binom{4}{2}=\frac{6}{2^4}\,.$$

E.g. after $n=5$ steps, the probability that the walk ends at $x=-1$, as shown in Figure 2, equals to:

$$p_{5,2}=\frac{1}{2^5}\binom{5}{2}=\frac{10}{2^5}\,.$$

Both of the binomial coefficient values in the numerator could be read directly from the above Pascal's triangle, and the total number of paths in the denominator can be found as the sum of the binomial coefficient in each triangle row.

3.2 Binomial probability distribution

A careful reader could have noticed that in the previous deliberation we have introduced the binomial probability distribution. In short, having a binary set of (elementary) events $\Omega_X=\{x_1,x_2\}$, with probabilities $P(x_1)=p$, $P(x_2)==q$, $p+q=1$, the probability for k-occurrences of the event x_1 (and $n-k$ occurrences of the event x_2) after n repetitions of the experiment is:

$$p_{n,k} = \binom{n}{k} p^k q^{n-k} \quad . \qquad (15)$$

In the first approximation we have assumed that the probabilities to go either left or right are equal, resulting in $p = q = 1/2$. Now the binomial distribution simplifies to:

$$p_{n,k} = \frac{1}{2^n}\binom{n}{k} \quad . \qquad (16)$$

In our case the event x_1 (x_2) corresponds to stepping to the left (right) and changing the distance x for the amount $\Delta x = -1$ ($\Delta x = +1$).

3.3 The mean (expected) values

To relate the distance x from the origin to the number n of random steps and the index k of the ending node, we write it as:

$$x = x(n,k) = x(n)_k\,, \qquad n = 0,1,\dots\dots \quad k = 0,1,\dots,n\,. \qquad (17)$$

We may say that to each node (n,k) we attribute the corresponding distance from the origin. From the discussion in 3.1 it can be easily concluded that:

$$x(n)_k = -n + 2k\,, \qquad k = 0,1,\dots,n\,. \qquad (18)$$

The extreme distances are:

$$x(n)_{min} = -n\,, \qquad k_{min} = 0\,, \qquad x(n)_{max} = n\,, \qquad k_{max} = n, \qquad (19)$$

and all the possible distances can be outlined as:

$$x(n) = \qquad (20)$$
$$= \begin{cases} -n, -n+2, \dots, -2, 0, 2, \dots n-2\,, n\ ; & n \text{ is even,} \\ -n, -n+2, \dots, -1, 1, \dots n-2\,, n\ ; & n \text{ is odd.}\end{cases}$$

From (18, 20) it is obvious that for an odd $n = 2l + 1$, $x(n)_k$ cannot be zero for any value of $l = 0, 1, \dots, \lceil n/2 \rceil$, which explicitly proves our earlier observation.

Having derived the probability and distance of the node (n,k) in equations (16) and (18), we are ready to write the expressions for the calculation of the main statistical indicators of the random walk. In general, for any value $r = r(n,k)$ being a function of the node parameters n and k, we calculate its mean (expected) value (see the discussion in 2.2) according to (3) as:

$$\langle r(n) \rangle = \sum_{k=1}^{n} p_{n,k} r(n,k). \qquad (21)$$

By substituting for $r(n,k)$ the values of $x(n)$, $|x(n)|$, $x^2(n)$, obtained from (18), we can summarize:

1. **The mean distance (MD)**

$$\langle x(n) \rangle = \frac{1}{2^n}\sum_{k=1}^{n}\binom{n}{k}(-n+2k)\,; \qquad (22.1)$$

2. **The mean absolute distance (MAD)**

$$\langle |x(n)| \rangle = \frac{1}{2^n}\sum_{k=1}^{n}\binom{n}{k}|-n+2k|\,; \qquad (22.2)$$

3. **The mean square distance (MSD)**

$$\langle x^2(n) \rangle = \frac{1}{2^n}\sum_{k=1}^{n}\binom{n}{k}(-n+2k)^2. \qquad (22.3)$$

The evaluation of the first and the third expression is straightforward, and will be given here for the sake of completeness. The second expression, presenting the central topic of the paper, is left for the next section.

In evaluating the (22.1) we obtain:

$$\langle x(n) \rangle = \frac{1}{2^n}\left[-n\sum_{k=1}^{n}\binom{n}{k} + 2\sum_{k=1}^{n}k\binom{n}{k}\right].$$

The first sum is a well known series amounting to 2^n, and the second can be obtained after a bit more elaborate calculation $\sum_{k=1}^{n} k\binom{n}{i} = n2^{n-1}$ (e.g. confer [7]). Now we conclude:

$$\langle x(n) \rangle = \tfrac{1}{2^n}\left[-n2^n + 2n2^{n-1}\right] = 0\,, \qquad (22.1')$$

which explicitly confirms the result in (7).

Similarly, the expression (22.3) can be rewritten as:

$$\langle x^2(n) \rangle = \frac{1}{2^n}\left[n^2\sum_{k=1}^{n}\binom{n}{k} - 4n\sum_{k=1}^{n}k\binom{n}{k} + 4\sum_{k=1}^{n}k^2\binom{n}{k}\right].$$

Here we use the results of the previous two series, together with $\sum_{k=1}^{n} k^2\binom{n}{i} = (n+n^2)2^{n-2}$ [7], wherefrom we get:

$$\begin{aligned}\langle x^2(n) \rangle &= \tfrac{1}{2^n}\left[n^2 2^n - 4n^2 2^{n-1} + 4(n+n^2)2^{n-2}\right] \\ &= \frac{1}{2^n}\left[n^2 2^n - 2n^2 2^n + (n+n^2)2^n\right] \\ &= n\,. \end{aligned} \qquad (22.3')$$

This directly proves the previous inductive derivation resulting in (12).

4. THE MEAN ABSOLUTE DISTANCE IN 1D RANDOM WALK

After having outlined the standard statistical parameters, in this section we come to the central topic of the paper –- the derivation of the mean absolute distance (MAD), or the mean absolute deviation of 1D random walk. Following the discussion in the previous section and the expression (21), the MAD value is:

$$\langle |x(n)| \rangle = \sum_{k} p_k |x(n)_k|\,, \qquad (23)$$

and is already formalized in the expression (22.2).

Before the evaluation, let's get a better insight in mean absolute distance by finding its values for the first few numbers of steps. With no steps taken $n = 0$ and we are still at the starting point:

$$\langle |x(0)| \rangle = 0\,.$$

After $n = 1$ step, the position is either $x(1) = -1$ or $x(1) = 1$:

$$\langle |x(1)| \rangle = \tfrac{1}{2}|-1| + \tfrac{1}{2}|1| = 1\,.$$

The mean value after $n = 2$ steps has three possible positions: −2, 0, 2. Moreover, there are two ways leading to the position $x = 0$, via -1 or $+1$ in the previous step. Thus we have:

$$\langle |x(2)| \rangle = \frac{1}{4}|-2| + \frac{2}{4}|0| + \frac{1}{4}|-2| = 1\,.$$

All the probability values can be directly followed in the Pascal's triangle in 2. The calculation of the mean values of the absolute distances for the first 5 steps is summarized in the following expression, and for the first 12 steps Table 1.

$$\langle|x(0)|\rangle = 0\,,$$
$$\langle|x(1)|\rangle = \tfrac{1}{2}(|-1| + |1|) = 1\,,$$
$$\langle|x(2)|\rangle = \frac{1}{2^2}(|-2| + 2|0| + |-2|) = 1\,,$$
$$\langle|x(3)|\rangle = \frac{1}{2^3}(|-3| + 3|-1| + 3|1| + |3|) = \frac{3}{2}\,,$$
$$\langle|x(4)|\rangle = \frac{1}{2^4}(|-4| + 4|-2| + 6|0| + 4|2| + |4|) = \frac{3}{2}\,,$$
$$\langle|x(5)|\rangle = \frac{1}{2^5}(|-5| + 5|-3| + 10|-1| + 10|1| + 5|3| + |5|) = \frac{15}{8}\,,$$
$$\dots\ \dots\ \dots$$

Table 1. The mean absolute distance for 12 steps. For every even step the mean absolute distance (MAD) is the same as for the previous even number of steps (if existing), which is designated as "– || –" . In the two rightmost columns the increase of the MAD values from the previous step is shown.

n	$\langle\|x(n)\|\rangle$	$\langle\|x(n)\|\rangle - \langle\|x(n-1)\|\rangle$	
0	0	–-	
1	1	1	$= 2^{0}\binom{0}{0}$
2	– || –	0	
3	3/2	1/2	$= 2^{-2}\binom{2}{1}$
4	– || –	0	
5	15/8 = 1.875	3/8	$= 2^{-4}\binom{4}{2}$
6	– || –	0	
7	35/16 ≈ 2.188	5/16	$= 2^{-6}\binom{6}{3}$
8	– || –	0	
9	315/128 ≈ 2.461	35/128	$= 2^{-8}\binom{8}{4}$
10	– || –	0	
11	693/256 ≈ 2.707	63/256	$= 2^{-10}\binom{10}{5}$
12	– || –	0	

From (25) and Table 1 we can see that only the odd steps contribute to the increase of the mean absolute distance, while the next even step leaves it at the same value. By inspecting the differences between the MAD values of the successive odd steps (see the third column in Table 1), it can be induced:

$$\langle|x(n)|\rangle - \langle|x(n-1)|\rangle = \frac{1}{2^{n-1}}\binom{n-1}{\frac{n-1}{2}},\ if\ \ n \text{ odd.} \quad (24)$$

In the derivation of this expression the Pascal's rule

$$\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}$$

should be used. The formal proof of this is out of the scope of this paper and will be presented elsewhere.

The equation (24) leads us to the following recursive formula:

$$\langle|x(n)|\rangle = \begin{cases} \langle|x(n-1)|\rangle + \frac{1}{2^{n-1}}\binom{n-1}{\frac{n-1}{2}}, n \text{ is odd,} \\ \langle|x(n-1)|\rangle\ , \qquad n \text{ is even.} \end{cases} \quad (25)$$

Nevertheless, in order to find a closed-form formula, we should calculate the sum (22.2), which can be rewritten:

$$\langle|x(n)|\rangle = \tfrac{1}{2^n}\textstyle\sum_{k=0}^{n}\binom{n}{k}|n-2k|. \quad (26)$$

Since the absolute value function is involved, we must track two possible cases:

$$\sum_{k=0}^{n}\binom{n}{k}|n-2k| = \sum_{k=0}^{n}\binom{n}{k}\cdot\begin{cases} n-2k, & \text{if } k \le n/2 \\ 2k-n, & \text{if } k > n/2\,. \end{cases}$$

The total number of summation terms depends on the parity of n . If the number of steps n is even, there will be an odd number of nodes and accordingly an odd number of terms, like, for example, when $n = 4$:

$$x(4) = (4 - 2\cdot k) \in \{-4,-2\} \cup \{0\} \cup \{2, 4\}.$$

If the number of steps is odd, there will be an even number of summation terms, like for $n = 5$:

$$x(5) \in \{-5,-3,-1\} \cup \{1, 3, 5\}.$$

As we have mentioned before, only the odd steps contribute to the increase of the mean absolute value, so that we shall continue by considering their case. Now the sum can be separated into two groups, the positive and the negative, which simplifies the calculation:

$$\sum_{k=0}^{n}\binom{n}{k}|n-2k| = \sum_{k=0}^{n}\binom{n}{k}\cdot\begin{cases} n-2k, if\ k \le n/2 \\ 2k-n, if\ k > n/2 \end{cases}$$
$$= \textstyle\sum_{k=0}^{\frac{n-1}{2}}\binom{n}{k}(n-2k) + \sum_{k=\frac{n+1}{2}}^{n}\binom{n}{k}(2k-n)$$
$$= \sum_{k=0}^{\frac{n-1}{2}} n\binom{n}{k} - \sum_{k=\frac{n+1}{2}}^{n} n\binom{n}{k} - 2\sum_{k=0}^{\frac{n-1}{2}} k\binom{n}{k} + 2\sum_{k=\frac{n+1}{2}}^{n} k\binom{n}{k}. \quad (27)$$

The first two terms cancel because of the symmetry of the binomial coefficients in the Pascal's triangle:

$$\binom{n}{k} = \binom{n}{n-k}.$$

Therefore we have:

$$\sum_{k=0}^{n}\binom{n}{k}|n-2k| = -2\sum_{k=0}^{\frac{n-1}{2}} k\binom{n}{k} + 2\sum_{k=\frac{n+1}{2}}^{n} k\binom{n}{k}$$

$$\langle|x(n)|\rangle = -\frac{2}{2^n}\left(\sum_{k=0}^{\frac{n-1}{2}} k\binom{n}{k} - \sum_{k=\frac{n+1}{2}}^{n} k\binom{n}{k}\right). \qquad (28)$$

For example, for 5 steps this formula gives:

$$\langle|x(5)|\rangle = -\frac{2}{2^5}(0+5+20-(30+20+5))$$
$$= \frac{2}{2^5}\cdot 30 = \frac{15}{8}\ .$$

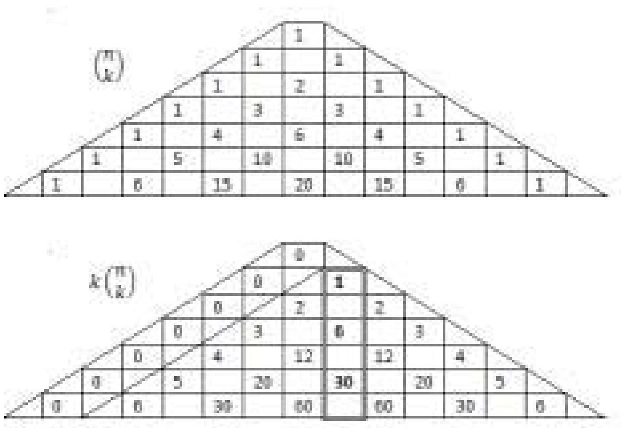

Figure 3. Symetrical subtriangle within the Pascal's triangle. When the binomial coefficients are multiplied with $\boldsymbol{k} = \boldsymbol{0}, \boldsymbol{1}, \ldots, \boldsymbol{n}$, a subtriangle appears within the main Pascal's triangle, with coefficients that are symmetrical to its central column in bold.

When the binomial coefficients $\binom{n}{k}$ are multiplied with k, an interesting thing happens: inside the Pascal's triangle, another, smaller triangle appears, which is symmetrical with respect to the 1st column to the right from the main triangle central column (the double border in the Figure 3). That's why all the terms cancel each other, except those inside it.

The first right column is associated with $k = (n+1)/2$ which leads us to the following closed-form expressions:

$$\langle|x(n)|\rangle = \frac{1}{2^{n-1}}\frac{n+1}{2}\binom{n}{\frac{n+1}{2}}, \quad n \text{ is odd}, \qquad (29a)$$

$$\langle|x(n)|\rangle = \frac{1}{2^{n-2}}\frac{n}{2}\binom{n-1}{n/2}, \qquad n \text{ is even}. \qquad (29b)$$

We shall prove the first formula for the odd number of steps by mathematical induction.

4.1 Proof by Induction

Let's start from the following statement for n is odd

$$\langle|x(i)|\rangle = \frac{1}{2^{i-1}}\frac{i+1}{2}\binom{i}{\frac{i+1}{2}}. \qquad (30)$$

Basis: $i = 1$

$$\langle|x(1)|\rangle = \frac{1}{2^0}\frac{1+1}{2}\binom{1}{\frac{1+1}{2}} = 1\ . \qquad (31)$$

Assumption: $i = n$

$$\langle|x(n)|\rangle = \frac{1}{2^{n-1}}\frac{n+1}{2}\binom{n}{\frac{n+1}{2}}\ . \qquad (32)$$

Inductive step: $i = n+2$.

If assumption holds, then it must be shown that:

$$\langle|x(n+2)|\rangle = \frac{1}{2^{n+1}}\frac{n+3}{2}\binom{n+2}{\frac{n+3}{2}}. \qquad (33)$$

From the recursive formula (25), it is:

$$\langle|x(n+2)|\rangle = \langle|x(n)|\rangle + \frac{1}{2^{n+1}}\binom{n+1}{\frac{n+1}{2}}. \qquad (34)$$

Now, according to the assumption in (32):

$\langle|x(n+2)|\rangle =$

$$= \frac{1}{2^{n-1}}\frac{n+1}{2}\binom{n}{\frac{n+1}{2}} + \frac{1}{2^{n+1}}\binom{n+1}{\frac{n+1}{2}} \qquad (35)$$

$$= \frac{1}{2^n}\frac{(n+1)!}{\left(\frac{n+1}{2}\right)!}\left[\frac{1}{\left(\frac{n-1}{2}\right)!} + \frac{1}{2\left(\frac{n+1}{2}\right)!}\right] \qquad (36)$$

$$= \frac{1}{2^n}\frac{(n+1)!}{\left(\frac{n+1}{2}\right)!}\left[\frac{\frac{n+1}{2}\cdot\frac{n+3}{2}}{\left(\frac{n+3}{2}\right)!} + \frac{\frac{n+3}{2}}{2\left(\frac{n+3}{2}\right)!}\right] \qquad (37)$$

$$= \frac{1}{2^n}\frac{(n+1)!\frac{1}{4}(n+3)}{\left(\frac{n+1}{2}\right)!}\cdot\frac{(n+2)}{\left(\frac{n+3}{2}\right)!}\quad . \qquad (38)$$

It can be easily shown that:

$$\langle|x(n+2)|\rangle = \frac{1}{2^{n+1}}\frac{n+3}{2}\binom{n+2}{\frac{n+3}{2}}. \qquad (39)$$

A very similar calculation can be done for the even number of steps.

4.2 Mean absolute value for a great number of steps

According to the Stirling's formula:

$$\lim_{n\to\infty}\frac{n!}{\sqrt{2\pi n}\left(\frac{n}{e}\right)^n} = 1\ , \qquad (40)$$

if n is replaced with $(n+1)/2$, it is

$$\lim_{n\to\infty}\frac{\left(\frac{n+1}{2}\right)!}{\sqrt{2\pi\frac{(n+1)}{2}}\left(\frac{n+1}{2e}\right)^{\frac{n+1}{2}}} = 1\ , \qquad (41)$$

as well as

$$\lim_{n\to\infty}\frac{\left(\frac{n-1}{2}\right)!}{\sqrt{2\pi\frac{(n-1)}{2}}\left(\frac{n-1}{2e}\right)^{\frac{n-1}{2}}} = 1. \qquad (42)$$

If we divide (40) by (41) and (42):

$$\lim_{n\to\infty} 2\pi \frac{\sqrt{\frac{(n+1)}{2}\frac{(n-1)}{2}}\left(\frac{n+1}{2e}\right)^{\frac{n}{2}+\frac{1}{2}}\binom{n}{\frac{n+1}{2}}}{\sqrt{2\pi n}\left(\frac{n}{e}\right)^{n}\left(\frac{n-1}{2e}\right)^{-\frac{n}{2}+\frac{1}{2}}} = 1$$

$$\lim_{n\to\infty} 2\pi \frac{\frac{(n+1)}{2}\left(\frac{n^2-1}{4e^2}\right)^{\frac{n}{2}}\binom{n}{\frac{n+1}{2}}}{\sqrt{2\pi n}\left(\frac{n}{e}\right)^{n}} = 1 \qquad (43)$$

$$\lim_{n\to\infty} 2\pi \frac{\frac{(n+1)}{2}\frac{1}{2^n}\left(\frac{n^2-1}{n^2}\right)^{\frac{n}{2}}\binom{n}{\frac{n+1}{2}}}{\sqrt{2\pi n}} = 1 \qquad (44)$$

Since $\lim_{n\to\infty}\left(1-\frac{1}{n^2}\right)^n = 1$, we can conclude that:

$$\sqrt{\frac{\pi}{2}}\lim_{n\to\infty}\frac{\frac{1}{2^{n-1}}\frac{(n+1)}{2}\binom{n}{\frac{n+1}{2}}}{\sqrt{n}} = 1 \qquad (45)$$

$$\lim_{n\to\infty}\langle|x(n)|\rangle = \sqrt{\frac{2n}{\pi}} \qquad (46)$$

This leads to the ratio between the mean absolute deviation and the standard deviation, in the limiting case:

$$\lim_{n\to\infty}\frac{\langle|x(n)|\rangle}{\langle x^2(n)\rangle} = \sqrt{\frac{2}{\pi}} \approx 0.7979\,. \qquad (47)$$

This ratio is the same as between MAD and SD of the Gaussian distribution. The reciprocal value is $\sqrt{\pi/2} \approx 1.2533$.

5. CONCLUSION

We have started our deliberation with a simple question why the intuitively simple notion of the mean absolute value is not used in statistical and physical analysis, and have exemplified this with the problem of 1D random walk. While the calculation of the mean squared values is straightforward, the one for mean absolute values encounters many complications and difficulties. As first, there is the disparity between the expressions for the odd and even numbers, that vanishes only in the limiting case of $n \to \infty$. The lengthy derivation and the complicated formula speak for themselves. The character of the binomial probability distribution remains in the final formula through the binomial coefficients. In a way, through the use of mean absolute values, we cannot get rid of the details of the process, which is in contrast the mean squared values. It is easy to conclude that the concepts that are easy to grasp, like the mean absolute deviation, may lead to the calculations and the results that are far from being neither easy to use nor elegant.

On the other hand, the difficulties involved in the derivations of the mean absolute values should not prevent us from using this "intuitive" notion. We have certainly shown that for 1D random walk one would think twice before building further on this approach. The elegance and simplicity of the final physical results and interpretations in broader context proved to be far more reaching in the history of science, than just the simple initial concepts, no matter how intuitive they may be.

Nevertheless, we believe that beyond the mere elegance and practicality, there are further, deeper, reasons for the use of the squared values and their averages. We find them throughout the physics, and also embedded into statistics. The usage of SD and RMS may have some physical background, besides their practical convenience and this ought to be explored. We hope to elaborate more on this subject in our future work.

6. REFERENCES

1. Feynman R., *Lectures on Physics*, §6 *Probability*, Addison-Wesley, USA, 1964.
2. Šips V., *Uvod u statističku fiziku*, *§4.9.Brownovo gibanje*, Školska knjiga, Zagreb, 1990.
3. Smith W.,Gillan M.J., *The Random Walk and the Mean Squared Displacement*, CCP5 Newsletter No.45, Daresbury, UK, 1996.
4. Gorard S. (2004.) *Revisiting a 90-year-old debate: the advantages of the mean deviation*, British Journal of Educational Studies, Vol. 53, Issue 4, 2005.
5. David H.A.(1998) *Early Sample Measures of Variability*, Statistical Science, Vol.13, No.4, 1998
6. Sheldon M. Ross, *Introduction to probability models* §2.4 *Expectation of a random variable*, Academic Press, USA, 2007.
7. http://ww.wikipedia.org, article Binomial coefficient, last access, June 2011.

Polytechnic of Varaždin,

J. Križanića 33, HR-42000 Varaždin, Croatia

jurica.hizak@velv.hr

robert.logozar@velv.hr