Bull. Aust. Math. Soc. **109** (2024), 471–475
doi:10.1017/S000497272300093X

ON THE N-POINT CORRELATION OF VAN DER CORPUT SEQUENCES

CHRISTIAN WEIß

(Received 20 June 2023; accepted 10 August 2023; first published online 15 September 2023)

Abstract

We derive an explicit formula for the N-point correlation $F_N(s)$ of the van der Corput sequence in base 2 for all $N \in \mathbb{N}$ and $s \geq 0$. The formula can be evaluated without explicit knowledge about the elements of the van der Corput sequence. This constitutes the first example of an exact closed-form expression of $F_N(s)$ for all $N \in \mathbb{N}$ and all $s \geq 0$ which does not require explicit knowledge about the involved sequence. Moreover, it can be immediately read off that $\lim_{N\to\infty} F_N(s)$ exists only for $0 \leq s \leq 1/2$.

2020 *Mathematics subject classification*: primary 11K16; secondary 11K38, 62H20.

Keywords and phrases: pair correlation statistic, van der Corput sequence, equidistribution.

1. Introduction

For a sequence $(x_n)_{n\in\mathbb{N}}$ of elements in $[0, 1]$, a natural number $N \in \mathbb{N}$ and $s \in \mathbb{R}_0^+$, we define

$$F_N(s) := \frac{1}{N}\#\Big\{1 \leq k \neq l \leq N : \|x_k - x_l\| \leq \frac{s}{N}\Big\}, \tag{1.1}$$

where $\|\cdot\|$ is the distance of a number from its nearest integer. It measures the behaviour of gaps between the first N elements of $(x_n)_{n\in\mathbb{N}}$ on a local scale. A sequence $(x_n)_{n\in\mathbb{N}}$ is said to have Poissonian pair correlations (see, for example, [2, 14]) if $F(s) := \lim_{N\to\infty} F_N(s) = 2s$ for all $s \geq 0$. As (1.1) corresponds to the N-point correlation in [6, 7] (with the f in the definition therein being chosen as the indicator function), we also call $F_N(s)$ the N-point correlation of $(x_n)_{n\in\mathbb{N}}$. To distinguish it from $F_N(s)$, we call $F(s)$ the limiting pair correlation function.

A generic uniformly distributed random sequence in $[0, 1]$ drawn from the uniform distribution has Poissonian pair correlations (see [8] for a proof). Nonetheless, there are few explicitly known such examples (see [2] and more recent examples in [5, 7]). One of the reasons why such examples are difficult to find is that it is, in general, hard to completely describe the gap structure of a finite sequence, that is, the lengths and combinatorics of gaps between neighbouring points (see [12]). Although research has mainly focused on the generic (Poissonian) case, the nongeneric case has also attracted more attention in recent times. In [9], it is shown that the limiting gap distribution of $(\{\log(n)\})_{n\in\mathbb{N}}$, where $\{\cdot\}$ denotes the fractional part of a number, has an explicit distribution which is not a Poissonian distribution but close to an exponential distribution. In [13], the limiting pair correlation function of $(\{\log(2n-1)/\log(2)\})_{n\in\mathbb{N}}$ is explicitly calculated by exploiting the simple gap structure of this sequence. Another result in [4] describes the limiting pair correlation function of orbits of a point in hyperbolic space under the action of a discrete subgroup. Finally, in [11], the nongeneric pair correlation statistic of the sequence $(n^\alpha)_{n\in\mathbb{N}}$ is studied for $0 < \alpha < 1$.

In this note, we add to the growing body of literature by calculating for all $N \in \mathbb{N}$ and $s \geq 0$ the N-point correlation $F_N(s)$ of the van der Corput sequence in base 2. The van der Corput sequence is a classical example of a low-discrepancy sequence and thus, in particular, a uniformly distributed sequence. Because of their importance in uniform distribution theory, their intuitive geometry and their generalisations to higher dimensions, van der Corput sequences are widely discussed in the literature (see [1, 3, 10, 14]).

Recall that for an integer $b \geq 2$, the b-ary representation of $n \in \mathbb{N}$ is $n = \sum_{j=0}^{\infty} e_j b^j$ with $0 \leq e_j = e_j(n) < b$. The radical-inverse function is defined by $g_b(n) = \sum_{j=0}^{\infty} e_j b^{-j-1}$ for all $n \in \mathbb{N}$ and the van der Corput sequence in base b is given by $x_i := g_b(i-1)$ for $i \geq 2$. For convenience, we add $x_1 = 0$ as the first element of a van der Corput sequence because it simplifies the presentation of results in our context.

THEOREM 1.1. *Let $N \in \mathbb{N}$ and $s \geq 0$. Let the 2-ary representation of N be $N = \sum_{j=0}^{M} e_j 2^j$ with the coefficients $e_0, e_1, \ldots, e_M \in \{0,1\}$. Then for the van der Corput sequence $(x_n)_{n\in\mathbb{N}}$ in base $b = 2$ we have*

$$F_N(s) = \frac{1}{N} \sum_{k=0}^{M} e_k \bigg(\Big\lfloor \frac{s}{N} 2^k \Big\rfloor + \sum_{l=k+1}^{N} e_l \cdot 2 \cdot \Big\lceil \frac{1}{2} \Big\lfloor \frac{s}{N} 2^{l+1} \Big\rfloor \Big\rceil \bigg) 2^{k+1}. \tag{1.2}$$

To the best of our knowledge, this constitutes the first example of an exact closed-form expression of $F_N(s)$ for all $N \in \mathbb{N}$ and all $s \geq 0$, where the right-hand side does not rely on explicit knowledge of the involved sequence. Moreover, the expression on the right-hand side is surprisingly simple. The formula is superior in terms of running time because the time needed to evaluate the N-point correlation from the definition grows quadratically in N, while the running time to compute the right-hand side of (1.2) only grows logarithmically. For example, it would be almost infeasible to calculate the N-point correlation for a given $s > 0$ and $N = 10^9$ on a standard computer via (1.1), while the evaluation of (1.2) takes less than a second.

The main step in proving Theorem 1.1 is to decompose the set

$$\Big\{ 1 \leq k \neq l \leq N : \|x_k - x_l\| \leq \frac{s}{N} \Big\}$$

into several subsets, where the indices of the elements of $(x_n)_{n\in\mathbb{N}}$ depend on powers of 2 instead of N. This idea goes back to [14], where the weak limiting pair correlation

https://doi.org/10.1017/S000497272300093X Published online by Cambridge University Press

function of van der Corput sequences was calculated. In principle, our proof technique could be applied to van der Corput sequences in arbitrary base but the expression on the right-hand side of (1.2) would be much longer and more complex. Therefore, we decided here to restrict to the case where a short formula can be given. The reason why this can only be done in base $b = 2$ is that there are only at most two different gap lengths for all $N \in \mathbb{N}$, while there are up to three different gap lengths for all bases $b \geq 3$.

Our formula has the following application.

COROLLARY 1.2. *The limit* $\lim_{N\to\infty} F_N(s)$ *exists if and only if* $0 \leq s \leq \frac{1}{2}$. *In this case,* $F(s) = 0$.

2. Proof of results

We prove our main result for the van der Corput sequence in base $b = 2$ by applying the same decomposition of (1.1) into subsets as in [14].

PROOF OF THEOREM 1.1. Let us write the N-point correlation as

$$N \cdot F_N(s) = \underbrace{\#\left\{\|x_i - x_j\| \leq \frac{s}{N} : 1 \leq i \neq j \leq 2^M\right\}}_{=:A(s,M,N)} + 2\underbrace{\#\left\{\|x_i - x_j\| \leq \frac{s}{N} : 1 \leq i \leq 2^M, 2^M + 1 \leq j \leq N\right\}}_{=:B(s,M,N)} + \underbrace{\#\left\{\|x_i - x_j\| \leq \frac{s}{N} : 2^M + 1 \leq i \neq j \leq N\right\}}_{=:C(s,M,N)}.$$

Since the set $A(s, M, N)$ consists of all points x_i which are numbers of the form $k/2^M$ with $0 \leq k < 2^M$, its magnitude can be calculated immediately as

$$\#A(s, M, N) = \left\lfloor \frac{s}{N} 2^M \right\rfloor \cdot 2^M \cdot 2 =: a(s, M, N).$$

In the definition of the set $B(s, M, N)$, the x_i are again of the form $k/2^M$ with $0 \leq k \leq 2^M$, while the x_j all have the form $l/2^{M+1}$ with odd l such that $1 \leq l < 2^{M+1}$ by the definition of van der Corput sequences. Hence, it follows that

$$\#B(s, M, N) = \left\lceil \frac{1}{2} \left\lfloor \frac{s}{N} 2^{M+1} \right\rfloor \right\rceil \cdot (N - 2^M) \cdot 2 =: b(s, M, N).$$

Thus, it only remains to calculate $C(s, M, N)$. To do that, we first see that $(x_j)_{j=2^M+1}^N$ is the van der Corput sequence $(x_j)_{j=1}^{N-2^M+1}$ translated to the right by $2^{-(M+1)}$. Note that $\|x_i - x_j\|$ is invariant under simultaneous translation of x_i and x_j. If we treat the simpler situation $N = 2^M + 2^k$ with $k < M$, then we can proceed inductively and apply the formula for $A(s, M, N)$ which yields

$$\#C(s, M, N) = \left\lfloor \frac{s}{N} 2^k \right\rfloor \cdot 2^k \cdot 2,$$

because the set of type B is empty here. In the general case,

$$C(s, M, N) = \sum e_k \cdot \Big(a(s, k, N) + 2 \cdot b\Big(s, k, N - \sum_{l=k+1}^{M} e_l 2^l\Big)\Big),$$

where the factor 2 appears because sets of type B are counted twice. Substituting the corresponding expressions for $A(\cdot)$ and $B(\cdot)$ yields a sum of the form

$$\sum_{k=1}^{N} e_k \Big(\left\lfloor \frac{s}{M} 2^k \right\rfloor 2^{k+1} + 4 \left\lceil \frac{1}{2} \left\lfloor \frac{s}{M} 2^{k+1} \right\rfloor \right\rceil \sum_{l=1}^{k-1} e_l 2^l \Big).$$

Collecting powers of 2 yields the formula on the right-hand side of (1.2). □

Having (1.2) at hand, it is not hard to calculate the limiting behaviour of $F_N(s)$.

Proof of Corollary 1.2. For $2^M \leq N < 2^{M+1}$, the 2-ary representation of $N \in \mathbb{N}$ is of the form $N = 2^M + \sum_{k=1}^{M} e_k 2^k$. Thus,

$$\left\lfloor \frac{s}{N} 2^{N+1} \right\rfloor = 0$$

for $0 \leq s \leq \frac{1}{2}$ and the limit is $F_N(s) = 0$ by Theorem 1.1. Now let $s \in [l, l+1)$ for some $l \in \mathbb{N}_0$. Then, we have $F_N(s) = 2l$ for $N = 2^M$, again by Theorem 1.1. If $s \in (1/2 + l, 1/2 + l + 1]$ for some $l \in \mathbb{N}_0$, we choose N big enough such that

$$\frac{2^{M+1}}{2^M + 1} \cdot s > 2.$$

Then $F_N(s) \geq 2^{M+1}/(2^M + 1)$ for all $N = 2^{M+k} + 2^k$ with $k \in \mathbb{N}$. Thus, $F_N(s)$ cannot converge for $s > \frac{1}{2}$. □

Acknowledgement

Research on this article was conducted during a stay at the Université de Montréal, whom the author would like to thank for their hospitality.

References

[1] M. Drmota and R. Tichy, *Sequences, Discrepancies and Applications*, Lecture Notes in Mathematics, 1651 (Springer, Berlin, 1997).

[2] D. El-Baz, J. Marklof and I. Vinogradov, 'The two-point correlation function of the fractional parts of $\sqrt{n}$ is Poisson', *Proc. Amer. Math. Soc.* **143**(7) (2015), 2815–2828.

[3] L. Kuipers and H. Niederreiter, *Uniform Distribution of Sequences* (Wiley, New York, 1974).

[4] C. Lutsko, 'Directions in orbits of geometrically finite hyperbolic subgroups', *Math. Proc. Cambridge Philos. Soc.* **171**(2) (2020), 277–316.

[5] C. Lutsko, A. Sourmelidis and N. Technau, ‘Pair correlation of the fractional parts of αn^{θ}’, Preprint, 2023, arXiv:2106.09800.
[6] C. Lutsko and N. Technau, ‘Correlations of the fractional parts of αn^{θ}’, Preprint, 2021, arXiv:2112.11524.
[7] C. Lutsko and N. Technau, ‘Full Poissonian local statistics of slowly growing sequences’, Preprint, 2022, arXiv:2206.07809.
[8] J. Marklof, ‘Distribution modulo one and Ratner’s theorem’, in: *Equidistribution in Number Theory, An Introduction*, NATO Science Series, 237 (eds. A. Granville and Z. Rudnick) (Springer, Dordrecht, 2007).
[9] J. Marklof and A. Strömbergsson, ‘Gaps between logs’, *Bull. Lond. Math. Soc.* **45**(6) (2013), 1267–1280.
[10] H. Niederreiter, *Random Number Generation and Quasi-Monte Carlo Methods*, CBMS-NSF Series in Applied Mathematics, 63 (SIAM, Philadelphia, PA, 1992).
[11] R. Sayous, ‘Effective pair correlations of fractional powers of integers’, Preprint, 2023, arXiv:2306.00793.
[12] C. Weiß, ‘Some connections between discrepancy, finite gap properties and pair correlations’, *Monatsh. Math.* **199** (2022), 909–927.
[13] C. Weiß, ‘An explicit non-Poissonian pair correlation function’, Preprint, 2023, arXiv:2304.14202.
[14] C. Weiß and T. Skill, ‘Sequences with almost Poissonian pair correlations’, *J. Number Theory* **236** (2022), 116–127.

CHRISTIAN WEIß, Ruhr West University of Applied Sciences,
Duisburger Str. 100, D-45479 Mülheim an der Ruhr, Germany
e-mail: christian.weiss@hs-ruhrwest.de

https://doi.org/10.1017/S000497272300093X Published online by Cambridge University Press