COMMUNICATIONS

DE LA FACULTÉ DES SCIENCES
DE L'UNIVERSITÉ D'ANKARA

Série $A_1$: Mathématique

TOME 25 ANNÉE 1976

# Sing Condition For Moving Averages

by

Maide ORUÇ and C. Güner OMAY

3

Faculté des Sciences de l'Université d'Ankara
Ankara, Turquie

Communications de la Faculté des Sciences
de l'Université d'Ankara



---

La Revue "Communications de la Faculté des Sciences de l'Université d'Ankara" est un organe de publication englobant toutes les disciplines scientifiques représentées à la Faculté.

La Revue, jusqu'a 1975 à l'exception des tomes I, II, III, était composée de trois séries:

Série A: Mathématique, Physique et Astronomie.

Série B: Chimie.

Série C: Sciences naturelles.

A partir de 1975 la Revue comprend sept séries:

Série $A_1$: Mathématique
Série $A_2$: Physique
Série $A_3$: Astronomie
Série B : Chimie
Série $C_1$: Géologie
Série $C_2$: Botanique
Série $C_3$: Zoologie

En principe, la Revue est réservée aux mémoires originaux des membres de la Faculté. Elle accepte cependant, dans la mesure de la place disponible, les communications des auteurs étrangers. Les langues allemande, anglaise et française sont admises indifféremment. Les articles devront être accompagnés d'un bref sommaire en langue turque.

Adres: Fen Fakültesi Tebliğler Dergisi Fen Fakültesi, Ankara, Turquie.

# Sing Condition For Moving Averages

Maide ORUÇ and C. Güner OMAY*

**SUMMARY**

It is important in many experimental sciences (like Astronomy) to find out whether elements of a series of given numbers are arranged at random or not. Generally it is assumed that accidental errors satisfy the Gaussian Law of errors. This Law, however, yields only one of the necessary condition for a series of numbers to be considered as accidental erros, and it is not sufficient; for, it concerns the only distribution of numbers and not their arrangement. Thus more necessary conditions are needed.

In this paper "SIGN SEQUENCE" conditions for moving averages with two, three, and four terms are established and theoretical results are confirmed with experimental resuslt.

## 1. INTRODUCTION

It is important in many experimental sciences to find out whether the elements of a series of a given numbers are arranged at random or not. Generally it is assumed that accidental errors satisfy the Gaussian Law of errors. This Law, however, yields only one of the necessary conditions for a series of numbers to be considered as accidental errors. But this condition is not sufficient, for it concerns only the distribution of numbers and not their arrangement.

Kermack and McKendrick [1] have pointed out that, in an infinitely long series of numbers chosen at random, the average length of a "Run" is 2,5. A "RUN" is either a sequence of increasing terms beginning with a minimum and ending with a maximum, or a sequence of decreasing terms beginning with a maximum and ending with a minimum. The length of a run is the number of the terms of which the run consists.

*This paper is the abstract of C.G. Omay's doctoral thesis.

Another necessary condition for accidental errors was established by Gleisberg [2]. If a sequence of consecutive terms having the same sign is called "SIGN SEQUENCE" and if the length of a sign sequence is the number of terms in the sequence, then it was proved that in an infinitely long series of accidental arrors the average length of a sign sequence is 2.

Another necessary condition for accidental errors is given by Gökmen [3]. It is shown that in a sequence with infilinite number of elements the average length of "GROUP SEQUENCE" is $\frac{g}{g-1}$, where g is the number of groups into which the elements of the sequence is divided.

If moving averages are taken, it is obvious that the three conditions can not be directly applied to the new sequence. It is interesting to find out the new form of the above mentioned conditions in this case.

The "RUN" condition was given by GLEISSBERG and "SIGN SEQUENCE" condition is established in this paper.

## 2. "SING SEQUENCE" CONDITION FOR A SERIES OF MOVING AVERAGES OF RANDOM NUMBERS.

Sign Sequence condition for a series of moving averages of random numbers, depends on the number of terms in the moving averages. For this reason siqn sequence condition will be given for two, three and four, term moving averages series.

### 2.1. General Theorem for Averages Length of Siqn Sequence Conditions.

Let us denote, the probability of at least two consecutive elements having the same siqn with P.

*Theorem :* In a series, with elements half of which are positive and the other half are negative, if the numbers of the elements in the series approach infinity the average length of "SIGN Sequences" reaches $\frac{1}{1-P}$ which has a standard deviation equal to $\frac{P}{1-P}$

*Proof:* First of all, let us calculate "i" in terms of P. It is obvious that,

$$\text{Prob.} \left\{ \begin{array}{l} \text{The length of any Sign} \\ \text{Sequence to be at least "i"} \end{array} \right\} = P^{i-1} \qquad (2.1.1)$$

Thus,

$$\text{Prob.} \left\{ \begin{array}{l} \text{The length of sign} \\ \text{Sequence to be "i"} \end{array} \right\} = P^{i-1}-P^{i} \qquad (2.1.2)$$

By means of above probability one can easily get the mean length and variance of sing sequence, when the number of terms in the series approaches infinity, as follows,

$$\bar{u} = \sum_{i=1}^{\infty} i \; (P^{i-1}-P^{i}) = \frac{1}{1-P} \qquad (2.1.3.)$$

$$\sigma^2 = \sum_{i=1}^{\infty} \left(i - \frac{1}{1-P}\right)^2 (P^{i-1}-P^{i}) = \frac{P}{(1-P)^2} \qquad (2.1.4)$$

The result of the above theorem does not depend on the numbers of terms in the moving averages; But for each case the value of "P" must be calcualted separately.

**2.2. Sing Sequence Condition for Moving Averages With Two Terms.**

*Theorem:* Mean length of sign sequence for moving avarages with two terms from series of random numbers is equal to 3 with variance 6 when the number of terms in the series approaches infinity.

*Proof:* Let us have a series of random numbers uniformly distributed in the interval from —H to + H

$$m_1, m_1, \ldots, m_N, \ldots \qquad (2.2.1)$$

From this series, let us form a series of moving averages with two terms.

$$a_1,\ a_2, \ldots,\ a_{N-1}, \ldots \tag{2.2.2}$$

The relationship with the elements of thwo series is as follows;

$$a_1 = \tfrac{1}{2}(m_1+m_2),\ a_2 = \tfrac{1}{2}(m_2+m_3), \ldots,\ a_i = \tfrac{1}{2}(m_i+m_{i+1}), \ldots \tag{2.2.3}$$

In general two consecutive terms can be written as follows.

$$X = \frac{x_1+x_2}{2}\ ,\ Y = \frac{x_2+x_3}{2} \tag{2.2.4}$$

where $x_i$'s are random numbers. Let us denote their distribution function by $F_i(x_i)$

From theory of probability one can write the following:

$$\text{Prob } (X \leqq 0, Y \leqq 0) = \iiint_G dF_1(x_1)\, dF_2(x_2)\, DF_3(x_3) \tag{2.2.5}$$

where

$$G: \left(\frac{x_1+x_2}{2} \leqq 0,\ \frac{x_2+x_3}{2} \leqq 0, -\infty \leq x_2 \leq +\infty\right) \tag{2.2.6}$$

Hence:

$$\text{Prob. } \{(X \cap Y) > 0\} = 2\int_{-\infty}^{+\infty} F_1(-x_2)\, F_3(-x_2)\, dF_2(x_2) \tag{2.2.7}$$

where $F_1(-x_2)$ and $F_3(-x_2)$ are probability of X and Y to be negative respectively. The above result is independent of the numer of terms in moving averages.

Now we can apply this general result to the specific case. In equation (2.2.3) $m_{i+1}$ is common element for $a_i$ and $a_{i+1}$. Let us denote this element by X. As we pointed out earlier $m_i$, $m_{i+1}$ and X are uniformly distributed in the interval from —H to +H. Therefore they have distribution function equal to $\frac{1}{2H}$

Hence

$$\text{Prob. } \{(a_i \leqq 0)\} = P\ \{(a_{i+1} \leqq 0)\} = \frac{H-x}{2H}\ . \tag{2.2.8}$$

If we substitute these results into the equation (2.2.7) we get the following.

$$\text{Prob.}\ \{(a_i \cap a_{i+1}) > 0\} = 2\int_{-H}^{+H} \frac{(H-x)^2}{4H^2}\,\frac{dx}{2H} = \frac{2}{3} \qquad (2.2.9)$$

which is equal to the value of P we defined previously.

Hence:

$$\bar{u} = \frac{1}{1-P} = 3,\ \sigma^2 = \frac{P}{(1-P)^2} = 6 \qquad (2.2.10)$$

**2.3. Sign Sequence Condition for the Moving Averages With Three Terms.**

*Theorem :* Mean length of sign sequence for the moving averages with three terms from a series of random numbers is equal to 48/13 with variance 1680/169 when the number of terms in the series goes to infinity.

*Proof:* Similar to the previous proof. Let us have a series of random numbers uniformly distributed in the interval from —H to+H.

$$m_1,\ m_2, \ldots,\ m_N, \ldots \qquad (2.3.1)$$

From this series, let us form a series of moving averages with three terms:

$$b_1,\ b_2, \ldots,\ b_{N-1}, \ldots \qquad (2.3.2)$$

where

$$b_i = \tfrac{1}{3}\,(m_i + m_{i+1} + m_{i+2}) \qquad (2.3.3)$$

let us calculate probability "P" in this case; Since in two consecutive terms, two terms are common. This common part will be designated by y. Thus with this notation, we get:

$$b_i = \frac{m_i + y}{3},\ b_{i+1} = \frac{y + m_{i+1}}{3} \qquad (2.3.4)$$

One can get the probabilities of $b_i$ and $b_{i+1}$ to be negative in terms of y as follows:

| *Intervals* | $P(b_i \leqq 0)$ | $P(b_{i+1} \leqq 0)$ | |
|---|---|---|---|
| $-2H \leqq y \leqq -H$ | 1 | 1 | |
| $-H < y < +H$ | $\frac{H-y}{2H}$ | $\frac{H-y}{2H}$ | (2.3.5) |
| $+H < y \leqq +2H$ | 0 | 0 | |

Now we will calculate the density function of y:

Since

$$f(m_i) = \frac{1}{2H} \qquad \text{for } -H \leqq m_i \leqq +H$$

and

$$y = m_{i+1} + m_{i+2}.$$

Thus we get,

$$g(y) = \int_{-\infty}^{+\infty} f(y-m_{i+1})\, f(m_{i+1})\, dm_{i+1} \tag{2.3.6}$$

Since following limitations must be statisfied,

$$\begin{aligned} -H &\leqq m_{i+1} \leqq +H \\ -H &\leqq y - m_{i+1} \leqq +H \end{aligned} \tag{2.3.7}$$

we get the following intervals for y respectively:

$$\begin{aligned} -2H &\leqq y \leqq 0 \\ 0 &< y < +2H. \end{aligned} \tag{2.3.8}$$

Hence density function for y will be as follows:

$$g(y) = \frac{1}{2H^2}\int_{-H}^{y+H} dm_{i+1} = \frac{2H+y}{4H^2} \quad \text{for} \quad -2H \leqq y \leqq 0$$

$$g(y) = \frac{1}{2H^2}\int_{y-H}^{+H} dm_{i+1} = \frac{2H-y}{4H^2} \quad \text{for} \quad 0<y\leqq+2H \tag{2.3.10}$$

If we substitute the above results in equation (2.2.6) we get the following:

$$p = \text{Prob.} \{(b_i \cap b_{i+1}) > O)\} = 2 \left\{ \int_{-2H}^{-H} \frac{2H+y}{4H^2} dy + \right.$$

$$\left. \int_{-H}^{0} \frac{(H-y)^2}{4H^2} \frac{2H+y}{4H^2} dy + \int_{0}^{+H} \frac{(H-y)^2}{4H^2} \frac{2H-y}{4H^2} dy \right\} \quad (2.3.11)$$

Then the result will be:

$$P = 35/48. \quad (2.3.12)$$

With this value one can calculate the mean length and variance of a sign sequence:

$$\bar{u}_3 = 48/13$$
$$\sigma_3^2 = 1680/169 \quad (2.3.13)$$

**2.4. Sign Sequence Condition for Moving Averages With Four Terms.**

*Theorem :* Mean length of sign sequence for moving averages with four terms from a series of random numbers is equal to 30/7 with variance 690/49 when the number of terms in the series goes to infinity.

*Proof :* Similar to the previous theorems we will form moving averages with four terms $\{c_i\}$ from a series of random numbers each of which has uniform distribution in the interwal from —H to +H.

In the series of moving averages two consecutive terms have three terms in common. Let us denote this common part by z. Similar to the provious cases we can write the followings:

| *Intervals* | $P\ \{c_i \leqq O\}$ | $P\ \{c_{i+1} = O\}$ | |
|---|---|---|---|
| $-3H \leqq z < -H$ | 1 | 1 | |
| $-H \leqq z \leqq +H$ | $\frac{H-z}{2H}$ | $\frac{H-z}{H}$ | (2.4.1) |
| $+H < z \leqq +3H$ | 0 | 0 | |

The common part z, is:

$$z = m_i + m_{i+1} + m_{i+2} \tag{2.4.2}$$

If we rewrite z like

$$z = m_i + y \tag{2.4.3}$$

where

$$y = m_{i+1} + m_{i+2} \tag{2.4.4}$$

each has density function

$$h(m_i) = \frac{1}{2H} \qquad \text{for} \quad -H \leqq m_i \leqq +H$$

$$k(y) = \begin{cases} \dfrac{2H+y}{4H^2} & \text{for} \quad -2H \leqq y \leqq 0 \\[2ex] \dfrac{2H-y}{2H^2} & \text{for} \quad 0 \leqq y \leqq +2H \end{cases} \tag{2.4.5}$$

Hence, generally f(z) is as follows:

$$f(z) = \int_{-\infty}^{+\infty} k(z\text{-}m_i)\ h(m_i)\ dm_i \tag{2.4.6}$$

Now, we have to determine the intervals for specific cases. First of all let us consider the following intervals:

$$\begin{aligned} -H &\leqq m_i \leqq +H \\ -2H &\leqq z\text{-}m_i \leqq 0 \end{aligned} \tag{2.4.7}$$

From the above intervals we get the following intervals for z:

$$\begin{aligned} -H &\leqq z \leqq +H \\ -3H &\leqq z \leqq -H \end{aligned} \tag{2.4.8}$$

Hence, density function for z becomes:

$$f(z) = \frac{1}{8H^3} \int_{z}^{-H} (2H+z+m_i)dm_i = \frac{3H^2-2Hz-z^2}{16H^3} \tag{2.4.9}$$

$$\text{for } -H \leqq z \leqq +H,\ 2H \leqq y \leqq 0$$

and

$$f(z) = \frac{1}{8H^3} \int_{-H}^{z+H} (2H+z+m_i)dm_i = \frac{9H^2+6Hz+z^2}{16H^3} \quad (2.4.10)$$

$$\text{for } -3H \leqq z \leqq -H,\ 2H \leqq y \leqq 0$$

Similarly from

$$\begin{aligned} -H &\leqq m_i \leqq +H \\ 0 &\leqq z-m_i \leqq +2H \end{aligned} \quad (2.4.11)$$

we get

$$\begin{aligned} -H &\leqq z \leqq +H \\ +H &\leqq z \leqq +3H \end{aligned} \quad (2.4.12)$$

Thus density function f(z) becomes:

$$f(z) = \frac{1}{8H^3} \int_{-H}^{z} (2H-z+m_i)dm_i = \frac{3H^2+2Hz-z^2}{16H^3} \quad (2.4.13)$$

$$\text{for } -H \leqq z \leqq +H,\ 0 \leqq y \leqq +2H$$

and

$$f(z) = \frac{1}{8H^3} \int_{z-2H}^{+H} (2H-z+m_i)dm_i = \frac{9H^2+2Hz-z^2}{16H^3} \quad (2.4.14)$$

$$\text{for } +H \leqq z \leqq 3H,\ 0 \leqq y \leqq +2H$$

Hence by equation (2.2.7) we get the value of "P" in this case:

$$P = 2 \left\{ \int_{-H}^{H} \frac{(H-z)^2}{4H^2} \cdot \frac{3H^2+2Hz-z^2}{16H^3} dz + \int_{-H}^{+H} \frac{(H-z)^2}{4H^3} \times \right.$$

$$\left. \times \frac{3H^2-2Hz-z^2}{16H^3} dz + \int_{-3H}^{-H} \frac{9H^2+6Hz+z^2}{16H^3} dz \right\} = 23/30 \quad (2.4.15)$$

By means of the above value of P we get:

$$\begin{aligned} \bar{u} &= 30/7 \\ \sigma^2 &= 690/49 \end{aligned} \quad (2.4.16)$$

If number of terms in moving averages are more than four, Central limit theorem can be used to get the distribution function of common part.

### 2.5. Generalization of Distribution Function.

In all proofs, we assumed that $m_i$'s have uniform distribution. It can be shown that all proofs are true for any symmetric distribution.

Proof will be given for moving averages with two terms. Suppose that $m_i$ has a distribution function $f(m_i)$ and,

$$f(m_i) = f(-m_i) \tag{2.5.1}$$

Let us form moving averages with two terms from a series of $m_i$'s. Typical term in moving averages series will be:

$$a_i = \frac{1}{2}(m_i + m_{i+1}) \tag{2.5.2}$$

As we did before let us denote the common part in two consecutive terms with x.

Thus,

$$\begin{aligned} P\{a_i \leqq 0\} &= P\{a_{i+1} \leqq 0\} \\ P\{a_i \leqq 0\} &= P\{m_i \leqq -m_{i+1}\} = P\{m_i \leqq -x\} \end{aligned} \tag{2.5.3}$$

Let us define $\lambda_x$ as follows

$$\lambda_x = \int_{-\infty}^{x} f(t)\, dt \tag{2.5.4}$$

Thus

$$P\{a_i \geqq 0\} = P\{m_i \geqq -x\} = \int_{-x}^{+\infty} f(t)\, dt = \lambda_x \tag{2.5.5}$$

Similarly

$$P\{a_{i+1} \leqq 0\} = \lambda_x \tag{2.5.6}$$

On the other hand;

$$\int_{-\infty}^{+\infty} \lambda^2_x f(x)\, dx = 1/3 \tag{2.5.7}$$

Thus

$$P = 2 \int_{-\infty}^{+\infty} \lambda^2_x f(x)\, dx = 2/3 \tag{2.5.8}$$

This is the same result we got before.

## 3. Comparison of Theoretical and Experimental Results.

By means of scientific subroutine packages, 3 000 random numbers were generated from standart normal distribution and uniform distribution. We formed moving averages with two, three, and four terms from above numbers.

Since random numbers were in the interval from 0 to 1 we assumed all the numbers which were less than 0,50 as negative and the ones more than 0,50 as positive.

Thus for each case we counted all the sign sequence and got the mean length and standart deviations for all series.

Theoretical and experimental results are shown in the following table:

| The number of terms in moving averages | Theoretical Distribution $\bar{u}_i$ | Uniform Distribution $\bar{u}_i$ | Normal Distribution $\bar{u}$. | Standart Deviation |
|---|---|---|---|---|
| 2 | 3.000 | 2.925 | 3.014 | 0.045 |
| 3 | 3.692 | 3.676 | 3.747 | 0.058 |
| 4 | 4.286 | 4.235 | 4.343 | 0.068 |

When two experimental results are compared with the theoretical result, one can easily see that results from uniform distribution are closer to the theoretical results than results from normal distribution. But both are within the limit of $\mp\ \sigma$.

## REFERENCES

[1] **Kermack, W.O.:** *"Tests for Randomness in a Series of Numerical Observation"* Proceeding of the Royal Society of Edinburg, Vol. LVII P. 228–240, 1937

[2] **Gleissberg, W.:** *"Bedingungen für die Anordnung Zufaelliger Fehler"* Revue de la Faculté des sciences de l'université d'İstanbul Ssrie A vol. XLX p. 107–126, 1947.

[3] **Gökmen, T.:** *"Rasat Serileri için Gurup dizisi Kriteryumu"*

[4] **Gnedenko, B.V.:** *The Theory of Probability* Chelsea Pub. Comp. 1967

[5] *Generation of Random Variate,* Subroutine Randu, Gauss.

## ÖZET

Bir çok uygulamalı bilimlerde (Astronomi gibi) verilen bir sayı dizisinin ögelerinin dizilişinin rastgele olup olmadığının saptanması önemlidir.

Genellikle rastgele hataların Gauss Kanununa uyduğu kabul edilir. Ancak bu kanun bir sayı dizisinin rastgele hatalardan oluştuğunu göstermek bakımından bir gerek koşul olup yeter koşul değildir. Çünkü adı geçen kanun sadece sayıların dağılımı ile ilgilidir, dizilişi ile ilgili değildir.Bu nedenle daha fazla gerek koşullara gereksinme vardır.

Bu makalede iki, üç ve dört terimli kayan ortalamalar dizisi için "İŞARET DİZİSİ" koşulu geliştirilmiştir. Geliştirilen bu teorik sonuçların deneysel sonuçlara uygunluğu da ayrıca kontrol edilmiştir.

**Prix de l'abonnement annuel**

Turquie: 15 TL; Étranger: 30 TL.
Prix de ce numéro: 5 TL (pour la vente en Turquie).
Prière de s'adresser pour l'abonnement à: Fen Fakültesi Dekanlığı Ankara, Turquie.

**Ankara Üniversitesi Basımevi. Ankara–1977**