V.M. Loznikov. International Journal of Engineering Research and Applications
www.ijera.com
ISSN: 2248-9622, Vol. 14, Issue 11, November 2024, pp 59-61

RESEARCH ARTICLE OPEN ACCESS

On the equivalence of the correlation function and the distribution of time intervals between all pairs of time series events.

V.M. Loznikov
FSBSI Space Research Institute of the Russian Academy of Sciences

Abstract
This note discusses the relationship between the distribution of $SI(k)$ time intervals between all pairs of time series events and the non-normalized biased estimate of the correlation function $r_{xx}(k)$. For a discrete series of events, the equivalence of $r_{xx}(k)$ and $SI(k)$ is proved in two ways (algebraic and "geometric").

Date of Submission: 10-11-2024 Date of acceptance: 22-11-2024

I. Introduction

There are two classical definitions for estimating the correlation function $R_{xx}(k)$. The first definition follows from the Wiener-Khinchin theorem [1, 2] for discrete time

$$R_{xx}(k) = \frac{1}{N}\sum_{n=0}^{N-1} P_{xx}(n)\, e^{i2\pi k\frac{n}{N}}, \qquad (1)$$

where $P_{xx}(n)$ is the density of the power spectrum of a random process $x(i)$, $\nu = \frac{n}{N\Delta T}$ is the frequency, k = 0, ..., (N-1) is the number of shifts on the time scale with discreteness ΔT, $N\Delta T$ is the length of the series.

The second definition [1, 2] uses the value of a random variable $x(i)$, i = 0, 1, ..., (N-1)

$$R_{xx}(k) = \frac{1}{N-|k|}\sum_{j=0}^{N-|k|-1} x(j+|k|)\, x(j)\,. \quad (2)$$

This unbiased estimate of $R_{xx}(k)$ is related to the biased estimate of $r_{xx}(k)$ by the simple line ratio $r_{xx}(k) = \frac{N-|k|}{N} R_{xx}(k)$. Note the fact that the coefficient $\frac{N-|k|}{N}$ has a triangular shape depending on k. We will need this in the future.

Distribution of intervals between events

The distribution of time intervals between successive random events is well known. For a Poisson series of random events, this is an exponential distribution $f_0(t) = \lambda\, e^{-\lambda t}$ [3].

The distribution of time intervals between pairs of random events with the omission of m events is also known. This is an m-order Erlang distribution. For a Poisson series of random events, it has the form $f_m(t) = \frac{\lambda(\lambda t)^m}{m!}\, e^{-\lambda t}$ [3].

We will be interested in the distribution of time intervals between all pairs of events [4, 5], i.e. the distribution of a random variable, which is the sum of all random variables having Erlang distributions $f_m(t)$ of all orders (m= 0, 1, ..., N).

Algebraic proof.

Let's give two discrete random processes $x(i)$ (i=0, 1..., n) and $y(j)$ (j=0, 1..., m) with a sampling bin ΔT. Let's **first assume** that $x(i)$ and $y(j)$ can only take the values **0 or 1**, and ***n*=*m*=*N***. We will call the value $z(i)$ an event if $z(i_z)$=1 for i=i_z, and the time corresponding to the event is denoted as $T_z(i_z) = i_z\, \Delta T$, with ($z$=$x$ or y).

For all pairs of events $x(i_x)$ and $y(j_y)$, we define the time interval between these events $(T_x(i_x) - T_y(i_y)) = (i_x - j_y)\, \Delta T = k\, \Delta T$ (where k is an integer).

Now define estimation of the cross distribution of intervals $SI(k)$ as the number of intervals with the length of $k\Delta T$ for all pairs { $x(i_x)$, $y(i_y)$ } of events:

$$SI(k) = \sum_{\mathrm{i_x}} \sum_{\mathrm{j_y}} \delta(|k\Delta T - (i_x - j_y)\Delta \mathrm{T}|) \qquad (3)$$

where $\delta(|z|) = \begin{pmatrix} 1, & z = 0 \\ 0, & z \neq 0 \end{pmatrix}$, $\sum_{i_x}$ и $\sum_{j_y}$ – sums for all events $x(i_x)$, $y(j_y)$.

Each pair of events separated by interval $k\Delta T$ gives unit contribute in the sum.

Without changing the result, formula (3) can be rewritten as

$$SI(k) = \sum_{i=0}^{n} \sum_{j=0}^{m} x(i)\, y(j)\, \delta(|k\Delta T - (i - j)\Delta \mathrm{T}|) \qquad (4)$$

where $x(i) = 1$ for event, when $i{=}i_x$, and $x(i) = 0$, otherwise; similarly, $y(j) = 1$ for events, when $j{=}j_y$, and $y(j) = 0$, otherwise.

Now the summation is performed for all values i and j, but the multipliers $x(i)$ and $y(j)$ highlight the events in the sum.

It is easy to see that this formula (4) is also valid in the case when there are several coincident (in time) events, or when $u[i_u] > 1$, $u = (x$ or $y)$.

Since the product of the quantities $x(i)y(j)$ has the same effect as $\delta(|\, k - (i - j)\,|)$, if $(i = j + k)$, then formula (4) can be written as

$$SI(k) = \sum_{i=0}^{n} \sum_{j=0}^{m} x(i)\, y(j)\, \delta(|k\Delta T - (i - j)\Delta \mathrm{T}|) = \sum x(j + k)\, y(j) \qquad (5)$$

where the summation is performed for all allowed values of j.

It follows from (5) and (2) that the estimate of the cross-distribution of $SI(k)$ intervals coincides with the non-normalized (and biased) estimate of the cross-correlation for real x and y

$$SI(k) = R_{xx}(k)\,(N - |k|) = \sum_{j=0}^{N-|k|-1} x(j + |k|)\; y(j). \quad (6)$$

This statement can be proved in another **“geometric" way**.

Imaging that row $y(j)$ ($j = 0, 1, ..., m$) is arranged under row $x(i)$ ($i = 0, 1, ..., n$) ($n = m = N$), and $x(i)$ is shifted relate to $y(j)$ by $lag = k\Delta T$, i.e. $j = i + k$. Let us match corresponding $x(i)$ with $y(j)$, which can assume (for beginning) only 0 or 1 values, and look for the coincidence.

If there are coincidences of events for some $j_y = i_x + k$, then for each from of these events coincidences $x(i_x)\, y(j_y) = 1$ and moreover all intervals $|\, i_x - j_y\,|\Delta T$ between the events $x(i_x)$ and $y(j_y)$ equals to $lag = k\Delta T$, because shift is equivalent to time interval between events.

The number of events coincidence { $\sum_{i=0}^{N-|k|-1} x(i)\; y(i + |k|)$ } for shift $k\Delta T$ is equal to number of intervals { $SI(k) = \sum_{\mathrm{i_x}} \sum_{\mathrm{j_y}} \delta(|k\Delta T - (i_x - j_y)\Delta \mathrm{T}|)$} between events.

If we now assume that some $x(i)$ or/and $y(j)$ can be >1, then the argument remains correct, since for such events it can be imagined that there are several coincident events.

II. Discussion and Conclusion

Thus, for a discrete series of events, the equivalence of estimating the distribution of time intervals between all pairs of events and estimating an unnormalized biased correlation function is proved. Apparently, calculating the distribution of SI(k) intervals can be useful in analyzing rare events when the event times or time intervals between consecutive events are known.

It is easy to see from (6) that for Poisson random series (when $x(\boldsymbol{i})$ and $y(\boldsymbol{j})$ are uncorrelated), the mathematical expectation of the distribution of intervals $SI(k)$ for the length of the interval $k\Delta T$ is

$$E\{SI(k)\} = \sum_{j=0}^{N-|k|-1} \overline{x}\; \overline{y} = (N - |k|)\, \overline{x}\; \overline{y}, \qquad (7)$$

where $\overline{x}$ and $\overline{y}$ are the average values of $x(i)$ ($i = 0, 1, ..., n$) and $y(j)$ ($j = 0, 1, ..., m$) ($n = m = N$) random series, respectively, $k = 0, 1, ..., N - 1$.

Therefore, **for Poisson random variables, mathematical expectation of the distribution of intervals $E\{SI(k)\}$ has a triangular shape.**

At $m \neq n$, the distribution of intervals will have a trapezoidal shape.

The variance of SI(k) values is equal to

$$D\{SI(k)\} = D\left\{ \sum_{j=0}^{N-|k|-1} x(j + |k|)\; y(j) \right\}$$

$$D\{SI(k)\} = (N - |k|)\,[\,D\{x\}D\{y\} + \overline{x}^2 D\{y\} + \overline{y}^2 D\{x\}\,] \qquad (8)$$

due to the independence of Poisson intervals.

REFERENCES

[1]. Marple, Jr., S.L., 1987, Digital Spectral Analysis with Applications, Englewood Cliffs, NJ, Prentice-Hall, Inc., 1987.
(Марпл.-мл. С.Л., 1990, Цифровой спектральный анализ и его приложения, Пер. с англ. Хабарова О.И. и к.ф.-м.н. Сидоровой Г.А., под ред. д.т.н. Рыжака И.С., Изд-во. Мир, Москва, 1990.)

[2]. Jenkins G.M., Watts D.G., 1969, Spectral Analysis and its Applications, Holden-Day, San Francisco, Cambridge, London, Amsterdam, 1969.
(Дженкинс Г., Ваттс Д., 1971, Спектральный Анализ и его Приложения, вып. 1, пер. с англ. Писаренко В.Ф. с предисл. Яглома А.М., Изд.-во “Мир”, Москва, 1971.)

[3]. Черушева Т.В., Зверовщикова Н.В., 2021, Теория Массового Обслуживания. Учебное пособие, Пенза: Изд-во ПГУ, 2021.
(Cherusheva T.V., Zverovshchikova N.V., 2021, Theory of Queuing. Teaching aid, Penza: Publishing House of PSU, 2021.)

[4]. Лозников В.М., 1991, Метод спектра интервалов, Препринт ИКИ АН СССР № 1726. (Loznikov V.M., 1991, Interval spectrum method, Preprint of the SRI of the USSR Academy of Sciences No. 1726.)

[5]. Лозников В.М., 1992, Кросс-спектр интервалов, Препринт ИКИ РАН № 1816. (Loznikov V.M., 1992, Cross-spectrum of intervals, IKI RAS Preprint No. 1816.)