Fluctuations in 1D point patterns

Rosalba Garcia-Millan,[1,*] George Cantwell,[2,†] Konstantinos Aronis,[3,‡] and Carlos Marcelo[4,§]

[1] *Imperial College London*
[2] *University of Michigan*
[3] *Johns Hopkins School of Medicine*
[4] *University of Chicago*

Random point patterns with similar apparent characteristics (such as density of points) may display very different behaviours when it comes to the way these points are distributed in space. Strongly negatively correlated points (e.g. due to repulsive forces) show a high degree of order, whereas strongly positively correlated points (e.g. due to attractive forces) show clustering and hence, a high degree of disorder. These two extreme cases define a spectrum where point patterns are classified according to their fluctuations from small fluctuations (order, *hyperuniformity*) to large fluctuations (disorder, clustering). Point patterns with weak correlations show the same trend in their fluctuations as the typical uncorrelated case and are said to display Poisson-like behaviour.

In this report, we show numerical results on different one-dimensional point patterns both computer generated and experimentally obtained (heartbeat recordings). Furthermore, we regard the case of anomalously suppressed fluctuations (*hyperuniformity*) as a case of special interest and we try to derive the underlying mechanisms that give rise to this behaviour. We study the Dirichlet distribution as a model that can be tuned to display any behaviour from a periodic signal to a clustered signal.

This report is a summary of early work that we have done during the Complex Systems Summer School 2018 organised by the Santa Fe Institute, New Mexico. Throughout the report we discuss a number of directions that we can take to further research about the questions that we pose here.

I. INTRODUCTION

In this project we study the fluctuations of one-dimensional point patters. The fluctuations may be characterised by the variance of the number of points in a window of observation. The question we are interested in is how do the fluctuations depend on the size of the window of observation. The window of observation has the same dimensionality as the space in which the point pattern is embedded and is characterised by the linear length R. For example, if the point pattern is in two dimensions, R may be the radius of the circle that is being considered to count the number of points.
In general, for a d-dimensional point pattern, we define $N(R)$ as the number of points in the window of observation characterised by the length R.The variance of the number of points is then

$$\mathrm{Var}(N(R)) = \left\langle N(R)^2 \right\rangle - \left\langle N(R) \right\rangle^2 \propto R^{\lambda}, \qquad (1)$$

where λ is the exponent of interest that contains information about the fluctuations in the point pattern as the window increases. We consider the following three cases,

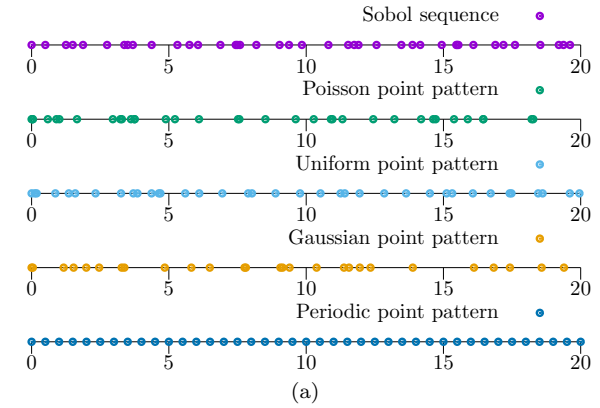
- if $\lambda > d$, the points in the pattern are heavily positively correlated, the pattern shows clustering (this case is related to Taylor's law [1]),
- if $\lambda = d$, the pattern is said to be Poisson-like, in the sense that the correlations are not strong enough to affect the uncorrelated typical behaviour (a Poisson process),
- if $\lambda \in [d-1, d)$, the points in the pattern are heavily anti-correlated (strong negative correlations), the pattern shows regularity (order) despite its randomness. The lower bound $d-1$ is given by the dimensionality of the surface of the window of observation, which is the minimum exponent the variance can scale with as R increases. This special case is known as *hyperuniformity* [2–4].

A. Numerics on 1D point patterns

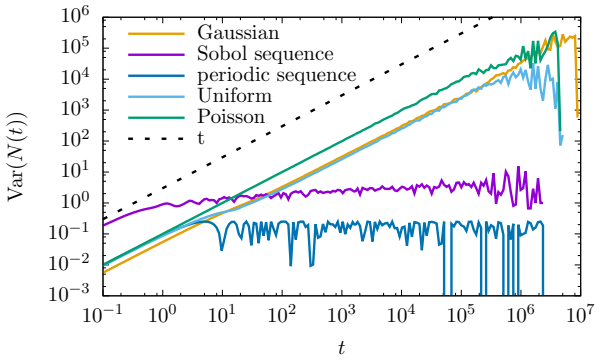
We have studied five point patterns numerically by drawing the waiting times (time between consecutive events) differently in each case: Poisson point pattern (exponential waiting times), uniformly distributed waiting times, normally distributed waiting times, the Sobol sequence, and a periodic point pattern. All of the cases considered have on average the same rate of occurrence (approximately 2 events per unit time) but the temporal correlations are different, see Fig. 1(a). To do the numerical analysis we have generated 10^6 waiting times in each case.

To estimate the curve $\mathrm{Var}(N(t))$ for each point pattern, we calculate the squared standard deviation of the number of points N in M different intervals of

* garciamillan16@imperial.ac.uk
† gcant@umich.edu
‡ karonis1@jhmi.edu
§ cmarceloservan@uchicago.edu



(a)



(b)

FIG. 1. (a) Sample of the considered 1D point patterns (b) Curves $\mathrm{Var}(N(t))$ for the different point patterns shown in Fig. 1(a), obtained from drawing 10^6 waiting times in each case and using Eq. (2). We observe that the point patterns "periodic" and "Sobol" show anomalous suppression of fluctuations (hyperuniformity) and "Gaussian", "uniform" and "Poisson" show the typical Poisson-like behaviour. The dashed line indicates the typical Poisson-like behaviour in 1D point patterns, namely the scaling form $\mathrm{Var}(N(t)) \propto t$.

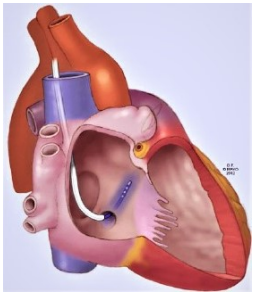
length t,

$$\mathrm{Var}(N(t)) = \frac{1}{M}\sum_{i=1}^{M}\left(\langle N^2\rangle_i - \langle N\rangle_i^2\right), \qquad (2)$$

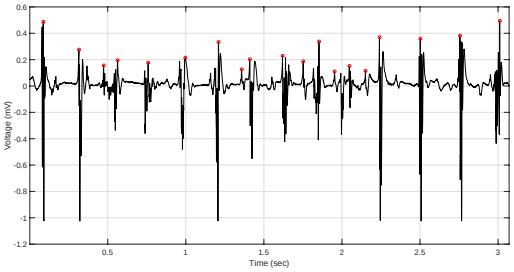
where $\langle\bullet\rangle$ denotes average.We then obtain curves like the ones shown in Fig. 1(b). As we expected, the Sobol sequence (see Sects. III B and II 0 d) and the periodic point pattern show anomalous suppression of fluctuations with exponent $\lambda = 0$. On the other hand, the "Gaussian", "uniform" and "Poisson" point patterns show the typical Poisson-like behaviour with exponent $\lambda = 1$.

B. Experimental data: heartbeat time series

We applied the same procedure to a point pattern generated from 40-minute recording of the cardiac activity of a patient with arrhythmia in five different locations of the heart. Fig. 2(a) shows a schematic cartoon of the experimental setup and Fig. 2(b) shows a sample of the recorded time series. The time series is mapped to a 1D point pattern as follows: the most pronounced spikes identified as the heartbeats are the events and the interspike intervals are the waiting times. See Sec. IV for more details.



(a)



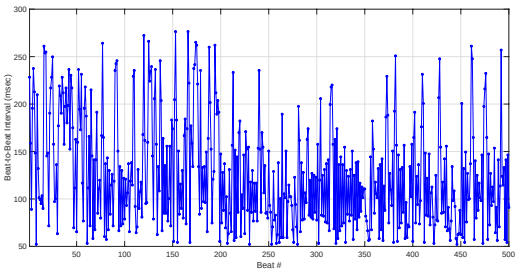
(b)

FIG. 2. In (a) we see a schematic cartoon of the position of the catheter inside the heart that measures and records the local electrical voltage in five different locations and in (b) we see an instance of the time series of measurements in a patient with arrhythmia, the red dots indicate those points (corresponding to suitable spikes in the data) in time where a heartbeat is considered to happen, cf. Sec. IV.

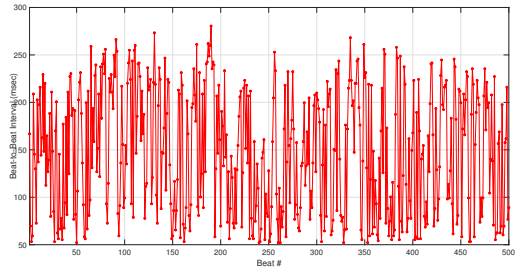
In Fig. 3(a) we show a sample of the interspike intervals as a function of heartbeat number and in Fig. 3(b) we show the same quantity for a simulated heartbeat time series. The simulations are generated so that they have the same frequency spectrum and autocorrelation structure as the experimental data.

In Fig. 4 we show our numerical results for the curves $\mathrm{Var}(N(t))$ for each of the five recordings and their corresponding surrogate time series. Our results show striking agreement in the trend of the fluctuations between the experimental data and the surrogate data. This a consequence of the fact that the surrogate data is generated so that it maintains the same autocorrelation structure as the experimental data because the behaviour that our observable $\mathrm{Var}(N(t))$ shows is based on the correlations between consecutive waiting times.

We find that there is a characteristic time scale around 200ms corresponding to the typical inter-peak interval length around which there is a change in the trend of



(a)



(b)

FIG. 3. These two graphs show the time intervals between spikes as a function of heartbeat number. In (a) we see experimental data from a patient with arrhythmia and in (b) surrogate data generated to simulate the experimental data in (a) keeping the same frequency spectrum and autocorrelation structure, cf. Sec. IV.

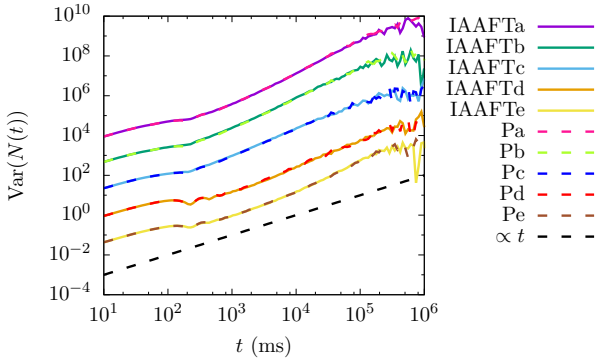


FIG. 4. Curves $\mathrm{Var}(N(t))$ for the five different time series obtained both from experimental recordings (Patient catheter a, b, c, d and e; dashed lines) and surrogate data (IAAFT a, b, c, d and e; solid lines), cf. Sec. IV. We observe that the fluctuations of the surrogate data is in good agreement with the fluctuations in the experimental data.

the fluctuations. For smaller times, the variance scales proportionally to the interval length ($\mathrm{Var}(N(t)) \propto t$), indicative of Poisson-like behaviour, and for larger times, the variance grows super-linearly, showing that the fluctuations increase faster and that the events are positively correlated, forming clusters.

II. INITIAL WORK AND ATTEMPTS DURING THE SUMMER SCHOOL

Initially, we explored the following question: Let (X_n) be a sequence of real random variables. Under which conditions do we have that $\mathbb{V}(\sum_i^n X_i) \sim n^\alpha$ for $\alpha < 1$? (as opposed to an independent identically distributed sequence for which the scaling is with $\alpha = 1$).

a. Random Walk To start with simple things, we tried a random walk $X_n = X_{n-1} + Y$ where $Y \sim \mathcal{N}(0,1)$. We showed that this process will always give a scaling with exponent 1.

b. Maximum entropy distribution One can formalize the scaling behavior as a constraint, and solve for the distribution that satisfies the constraint and has maximum entropy. Unfortunately this process seemed too cumbersome to lead to any analytical result (Especially because we are interested in the scaling with n, and we would need to solve for a distribution at each length n and look for the convergence of those distributions as $n \to \infty$).

c. Markov process We also tried to generate the sequence from a Markov process. We showed that for any two-state Markov process it is impossible to achieve the desired scaling.

d. Quasirandom numbers We found that apparently there is an interesting connection between the supressing of fluctuations and the quasi-Monte Carlo method to perform numerical integration. In that method, instead of using a lattice or sequence of random numbers in which to evaluate the function, a sequence of "*subrandom numbers*" is used. This sequence is characterized by a quantity called *discrepancy* and is also known as *low discrepancy sequence*. Our numerical experiments suggest that this type of sequences display the suppresion of fluctuations that we were looking for, therefore an interesting future direction is to either prove that they are indeed related or disprove it.

III. MORE ON HYPERUNIFORMITY

A. A model for hyperuniformity

To study any phenomenon, a model is useful. For "hyperuniformity", we propose using the symmetric Dirichlet distribution as a model. The PDF for the n-dimensional symmetric Dirichlet distribution is

$$P(\boldsymbol{x}|\alpha) = \frac{\Gamma(n\alpha)}{\Gamma(\alpha)^n} \prod_{i=1}^{n} x_i^{\alpha-1}. \tag{3}$$

The distribution is defined on the $(n-1)$-dimensional simplex, which means that all $x_i > 0$ and $\sum_i x_i = 1$. Our model for point patterns is the partial sums of

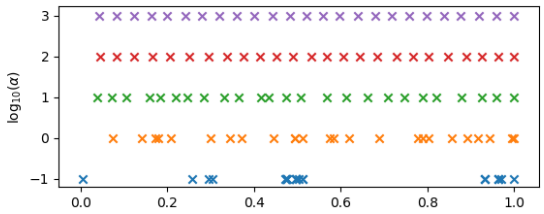


FIG. 5. Distribution of points, y_i, for different values of α.

Dirichlet distributed data, i.e.

$$y_j = \sum_{i=1}^{j} x_i. \tag{4}$$

The parameter α controls how lattice like the points y_i will be. When $\alpha = 1$, the y_i will be uniformly distributed within the unit interval. Otherwise, when α is very large, the y_i will fall on a lattice, and when α is small the y_i will be clustered together. Thus, this model has a wide range of behavior and allows us to continuously transition between points that are distributed uniformly and points on a lattice. See Fig. 5 for examples with different values of α.

When drawing from the model, $\alpha^{-1} = 0$ creates perfect lattices, while $\alpha^{-1} = 1$ is the uniform distribution. This allows us to use α^{-1} as a parameter to characterize the fluctuations, with $0 < \alpha^{-1} < 1$ corresponding to hyperuniformity. However, when we are presented with real-world data we aren't given a value of α. Thus to analyze data we will estimate a value for α and we can then use that to characterize the data. To estimate a value for α from we propose using the method of maximum likelihood. In other words, we maximize Eq. 3 with respect to α, or equivalently its logarithm:

$$\ln P(\boldsymbol{x}|\alpha) = \ln\Gamma(n\alpha) - n\ln\Gamma(\alpha) + (\alpha - 1)\sum_{i=1}^{n}\ln(x_i). \tag{5}$$

The maxima of Eq. 5 is the value of α that solves

$$\frac{\partial \ln P(\boldsymbol{x}|\alpha)}{\partial\alpha} = n\Psi(n\alpha) - n\Psi(\alpha) + \sum_{i=1}^{n}\ln(x_i) = 0, \tag{6}$$

Where Ψ is the digamma function. In general, a closed form solution to this equation does not exist. In practise, however, Newton's method converges quickly. So, noting that

$$\frac{\partial^2 \ln P(\boldsymbol{x}|\alpha)}{\partial\alpha^2} = n^2\Psi'(n\alpha) - n\Psi'(\alpha) \tag{7}$$

we iterate

$$\alpha^{(t+1)} = \alpha^{(t)} - \frac{\Psi(n\alpha^{(t)}) - \Psi(\alpha^{(t)}) + \frac{1}{n}\sum_{i=1}^{n}\ln(x_i)}{n\Psi'(n\alpha^{(t)}) - \Psi'(\alpha^{(t)})} \tag{8}$$

to convergence, starting with $\alpha^{(0)} = 1$.

B. Relation to integration

Conceptually, hyperuniformity seems to be related to low-discrepancy sequences. One of the useful applications of low-discrepancy sequences has been the quasi-Monte Carlo method. Similar to the regular Monte Carlo method, the quasi-Monte Carlo method is a numerical integration technique that scales more favorably in high dimensions than lattice techniques. However, the low-discrepancy property improves upon regular Monte Carlo methods by covering the space more evenly. We hypothesize that biological integrators (such as the retina) may use hyperuniformity to improve estimation in much the same way as numerical integration software uses low-discrepancy sequences.

IV. EXPERIMENTAL DATA

A. Methods of Experimental Design:

We collected data from a patient who was referred for a standard catheter ablation for symptomatic, drug-refractory AF at the Johns Hopkins Hospital in March 201. The protocol was approved by the Johns Hopkins Medicine Institutional Review Board and the participant provided written informed consent. The patient underwent pre-procedural transesophageal echocardiogram to rule out intracardiac thrombus. We introduced a 5-Fr decapolar catheter (Dynamic Tip 2-5-2 Boston Scientific, Marlborough, MA; inter-electrode distance 2mm between poles and 5mm between bipolar pairs) in the right femoral vein and advanced it to the coronary sinus. We recorded intracardiac bipolar electrograms from the decapolar catheter during AF using all 5 pairs of immediately adjacent electrodes. Sampling frequency was 977Hz and a duration of the recording was 40 minutes. We used a standard clinical electrophysiology recording system (CardioLab, GE Healthcare, Waukesha, WI). We defined the atrial depolarization as any peak exceeding 0.02mV in amplitude located at least 100 ms away from the prior peak, which was shorter than the atrial effective refractory period in all cases. We confirmed the accurate identification of atrial depolarization with visual inspection of all detected peaks. Iterative amplitude-adjusted Fourier transform surrogate data are generated such that the surrogate data have the same power spectrum, autocorrelations, and probability distribution with the original time series. As

a result, the derived surrogates have the same probability distribution and power spectrum with potential high order correlations being randomized. In the present study, we used IAAFT surrogate data created with the improved algorithm of Schreiber and Schmitz, [5] using the MathWorks implementation by Leontitsis et al.[6].

B. Results:

- Fluctuations of interpeak intervals of atrial electrical activity of a patient in atrial fibrillation exhibit Taylor's power law scaling.
- IAAFT surrogate data exhibited identical fluctuation profile with the original data. This is expected since, by design, the IAAFT surrogate data have the same linear autocorrelation profile with the original data.

C. Future plans:

- Obtain intracardiac recordings of similar length (40 mins) from patients that undergo catheter ablation for atrial fibrillation after normal sinus rhythm has been restored.
- Compare fluctuation patterns of intracardiac recordings in patients with atrial fibrillation before and after a successful therapeutic procedure.

ACKNOWLEDGMENTS

We would like to thank the Santa Fe Institute for organising the Complex Systems Summer School 2018 and our home institutions for making our attendance to the summer school possible.

[1] Z. Eisler, I. Bartos, and J. Kertész, Advances in Physics **57**, 89 (2008).
[2] A. Gabrielli, M. Joyce, and F. Sylos Labini, Phys. Rev. D **65**, 083523 (2002).
[3] S. Torquato and F. H. Stillinger, Phys. Rev. E **68**, 041113 (2003).
[4] D. Hexner and D. Levine, Phys. Rev. Lett. **118**, 020601 (2017).
[5] T. Schreiber and A. Schmitz, Phys. Rev. Lett. **77**, 635 (1996).
[6] `www.mathworks.com/matlabcentral/fileexchange/1597`.