Communications in Nonlinear Science and Numerical Simulation 120 (2023) 107166

Contents lists available at ScienceDirect

Communications in Nonlinear Science and Numerical Simulation

journal homepage: www.elsevier.com/locate/cnsns

Research paper

Periodic measures and Wasserstein distance for analysing periodicity of time series datasets

Chunrong Feng [a], Yujia Liu [b], Huaizhong Zhao [a,b,*]

[a] *Department of Mathematical Sciences, Durham University, DH1 3LE, United Kingdom*
[b] *Research Centre for Mathematics and Interdisciplinary Sciences, Shandong University, Qingdao 266237, China*

ARTICLE INFO

Article history:
Received 3 October 2022
Received in revised form 7 February 2023
Accepted 8 February 2023
Available online 11 February 2023

MSC:
primary 60B12
62M05
secondary 37A50
37A44

Keywords:
Random periodic processes
Ergodicity
Periodic measures
Law of large numbers
Dataset
Wasserstein distance

ABSTRACT

In this article, we establish the probability foundation of the periodic measure approach in analysing periodicity of a dataset. It is based on recent work of random periodic processes. While random periodic paths provide a pathwise model for time series datasets with a periodic pattern, their law is a periodic measure and gives a statistical description and the ergodic theory offers a scope of statistical analysis. The connection of a sample path and the periodic measure is revealed in the law of large numbers (LLN). We prove first the period is actually a deterministic number and then for discrete processes, Bézout's identity comes in naturally in the LLN along an arithmetic sequence of an arbitrary increment. The limit is a periodic measure whose period is equal to the greatest common divisor between the test period and the true period of the random periodic process. This leads to a new scheme of detecting random periodicity of a dataset and finding its period, as an alternative to the Discrete Fourier Transformation (DFT) and periodogram approach. We find that in some situations, the classical method does not work robustly, but the new one can work efficiently. We prove that the periodicity is quantified by the Wasserstein distance, in which the convergence of empirical distributions is established.

1. Introduction

Random periodicity is ubiquitous in the real world. It can be found, e.g., in daily temperature variations, economic cycles, internet traffic volume, the activity of sunspots, the EI Nino phenomenon and Earth's ice age transitions between "cold" and "warm" climates. Efforts for searching for periodic components or repeated patterns from data have been made for thousands of years including early Egyptian and Greek astronomers' observations of apparent periodic motion of the sun and the repeated patterns of sunspots. The early observation led to the establishment of the Julian calender. The periodogram, based on Fourier analysis of data, was introduced by [1] and has been used to analyse many datasets. Spectrum estimates and fast Fourier transformation have taken advantages of computer advances to carry out extensive computations. However, we note in this paper that the periodogram or Fourier analysis approach may fail to work robustly especially when the periodicity of the mean is weak or when a deterministic periodic function cannot approximate the time series. This can be the case when the noise is not stationary, but may have some periodic pattern. The Fourier

* Corresponding author at: Department of Mathematical Sciences, Durham University, DH1 3LE, United Kingdom.
E-mail addresses: chunrong.feng@durham.ac.uk (C. Feng), yujia.liu@sdu.edu.cn (Y. Liu), huaizhong.zhao@durham.ac.uk (H. Zhao).

https://doi.org/10.1016/j.cnsns.2023.107166

transformation method does not respond to the periodicity of the noise concerned. But the pattern of volatility is also important for applications in many areas such as in finance and option trading. Note a time series can be viewed as a (random) dynamical system which may be hidden. With this in mind, in this paper we provide an alternative novel approach to analyse periodicity in a dataset using periodic measures and their ergodic theory that we have been developing in the last fifteen years. The ergodic periodic measure (EPM) scheme works effectively in these situations and has a clear advantage to be able to detect periodicity of distributions. The equivalence of ergodicity and the law of large numbers suggests a way to the estimate of empirical periodic measures and the Wasserstein distance provides a quantitative approach to verify the periodicity of empirical periodic measures from datasets. Our aim in this paper is to give a probability theoretical account.

The concept of random periodic processes describes randomness and periodicity in the evolution of the stochastic processes simultaneously. The pathwise random periodic paths of random dynamical systems was first introduced in [2]. Later, the concept of random periodicity for semi-flows of random dynamical systems was established in [3,4]. In [3], the authors studied periodic measures which describe how the distribution of a random periodic process evolves periodically in time. They proved that the random periodic path and periodic measure are "equivalent" in some sense. They also obtained for the first time the ergodicity of periodic measures of the transition probability semigroups for Markovian systems. This result suggests that while random periodic processes provide a pathwise model for time series datasets with periodic pattern, periodic measures give a statistical description of the random periodicity. In [5], the authors defined random quasi-periodic paths for random dynamical systems and quasi-periodic measures for Markovian semigroups, which may provide a tool to study quasi-periodic phenomena in real life.

It is worth to mention here that the relevance of the random periodic paths, periodic measures and their ergodic theory to theoretical and applied problems arising in stochastic dynamical systems has began to be noted. In particular, there has been progress in the study of random periodicity on some topics e.g. bifurcations [6], random attractors [7], stochastic resonance [8–11], random horseshoes [12], modelling El Niño phenomenon [13], stochastic oscillations [14], large deviations [15], linear response and homogenisations [16,17], random almost periodic solutions [18,19], random periodic solutions of certain functional differential equations [20] and certain stochastic differential equations and stochastic partial differential equations [21–23].

Time series, which appears as a time evolution process in a pathwise fashion, does not give us how its law evolves along the time variable immediately. The ergodic theory of random periodic processes provide a perfect connection with the evolution of its law as a periodic measure. In particular, the law of large numbers (LLN) from the framework of Birkhoff's ergodic theorem [24] provides a scope for statistical analysis. This inspires us to establish a theory of time series analysis to help describing periodic phenomena of datasets. One of the challenges in the situation with uncertainty is to find the period which could be marred by random perturbations in a time series. In fact, at the first sight of a time series, the period may even be seen as random sometimes. In this paper, we prove a result which says that the period in the definition of a random periodic path is actually a deterministic number if the underlying noise metric dynamical system is ergodic. This justifies why in the classical Fourier series approach of time series and in our definition of random periodic paths, the periods of these random evolution processes are actually deterministic and certain patterns repeat with deterministic repetition times. On the other hand, when we look at the distributions, they are periodic with a deterministic period. This long over due result also justifies the same basic assumption in the new approach introduced in this paper.

The law of large numbers for a arithmetic progression with the common increment the same as the period follows from the definition of random periodic paths and Birkhoff ergodic theorem immediately under the assumption that the underlying noise metric dynamical system is ergodic. However, the period, even we have proved it is deterministic, may not be known to us in applications, e.g. in a time series dataset. To overcome this problem, we develop in this paper a law of large numbers along an arithmetic progression with an arbitrary common increment as a test period. Utilising Bézout's identity in number theory we prove the LLN along the sequence of arithmetic sequence with a common increment p, whose greatest common factor with the period of the periodic measure is the period obtained in the limit of the law of large numbers. Our convergence is proved in the Wasserstein distance. This result enables us to establish a new scheme to compute the true period of the random periodic process.

We would like to point out that, though in the proof of the LLN we use the theory of random dynamical systems, our final theorem is presented in a manner with little knowledge required on random dynamical systems and skew product dynamical systems. Thus it is convenient for application in statistics and time series analysis.

For a given time series $\{y_0, y_1, y_2, \ldots\}$, the LLN of arithmetic progression tells us the map: $i \mapsto \mu^{i,p}$, where $\mu^{i,p}$ is the empirical limit, $\frac{1}{K}\sum_{k=0}^{K-1}\delta_{y_{i+kp}}(\cdot) \to \mu^{i,p}(\cdot)$ as $K \to \infty$, has period r, where r is the greatest common divisor between p and q. In practice, the amount of data, even in the era of big data, is always limited. Thus we can only obtain approximations $\mu_K^{i,p}$ to the real periodic measures. The Wasserstein distance $W_1(\mu^{1,p}, \mu^{i,p})$, $i = 1, \ldots, r, \ldots$, gives a good way to quantify the periodicity of the map $i \mapsto \mu^{i,p}$, and the real valued function $i \mapsto W_1(\mu^{1,p}, \mu^{i,p})$ should be a periodic function of period r if $\mu^{i,p}$ is a periodic measure of period r. Needless to say, in practice, the periodicity of the map $i \mapsto W_1(\mu^{1,p}, \mu^{i,p})$ can be obtained approximately only by using empirical approximation of periodic measures $\mu^{i,p}$ as proved in Corollary 2.21. Note no priori information about the real period q is needed in the computation of the periodicity of $\mu^{i,p}$. But if the period of $\mu^{i,p}$ is $r \in \mathbb{N}^+$, then r is a factor of q. Using this scheme we can find all factors of q.

2. The law of large numbers

2.1. Random periodic paths and generic non-randomness of the period

Consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a Polish space $\mathbb{X}$, let $\Phi : \mathbb{R}^+ \times \Omega \times \mathbb{X} \to \mathbb{X}$ be a random dynamical system cocycle over a metric dynamical system $(\Omega, \mathcal{F}, \mathbb{P}; (\theta_t)_{t\in\mathbb{R}})$. To be more specific, as usual, we assume $\theta : \mathbb{R} \times \Omega \to \Omega$ is measurable with respect to $\mathcal{B}(\mathbb{R}) \otimes \mathcal{F}/\mathcal{F}$, θ_t preserves $\mathbb{P}$ for each $t \in \mathbb{R}$ and satisfies $\theta_t \circ \theta_s = \theta_{t+s}$ for all $t, s \in \mathbb{R}$. The space $\mathbb{X}$ is the state space where our processes lie in and is assumed to be a separable Banach space. For many statistical applications, it is adequate to take $\mathbb{X} = \mathbb{R}^d$ with the Euclidean norm. The random map Φ evolving on the state space when time moves satisfies: (i) The map Φ is $\mathcal{B}(\mathbb{R}^+) \otimes \mathcal{F} \otimes \mathcal{B}(\mathbb{X})$-measurable; (ii) $\Phi(0, \omega) = I$ for a.e. $\omega \in \Omega$; (iii) $\Phi(t+s, \omega) = \Phi(t, \theta_s\omega) \circ \Phi(s, \omega)$, $t, s \in \mathbb{R}^+$ for a.e. $\omega \in \Omega$.

The definition of random periodic path of random dynamical system Φ is given in [2,4]. See also [3].

Definition 2.1. A random periodic path of period T of the random dynamical system Φ: $\mathbb{R}^+ \times \Omega \times \mathbb{X} \to \mathbb{X}$ is an $\mathcal{F}$-measurable map $Y : \mathbb{R} \times \Omega \to \mathbb{X}$ such that for almost all $\omega \in \Omega$,

$$\Phi(t, \theta_s\omega)Y(s, \omega) = Y(t + s, \omega) \tag{2.1}$$

and for any $t \in \mathbb{R}^+$, $s \in \mathbb{R}$,

$$Y(s + T, \omega) = Y(s, \theta_T\omega). \tag{2.2}$$

Remark 2.2. (i) For a statistical description, we usually do not know the exact expression of the dynamical system driving the time series. But it is hidden in the time series evolution. In fact a time series can be regarded as a (random) dynamical system. The theoretical existence of Φ in a time series helps us to use advances in the study of random periodic processes and periodic measures to establish a time series theorem.

(ii) Many people will have expected that the period of a random periodic path might be random rather than deterministic. Note in Definition 2.1, as a basic assumption of this paper, the period T is a deterministic number rather than a random variable. This is probably against intuitive instincts in the first sight. However, there are many reasons for that. This makes sense with the help of the metric dynamical system of noise in the definition of the random periodic paths in the pathwise sense. In fact, it is not expected that the random periodic path will come back to the same position after one period, sometimes even not somewhere near. This "occasionally nowhere near" feature is allowed in Definition 2.1 of random periodic path. Nevertheless, it was proved in [3] that the definitions of random periodic paths and periodic measures (see Definition 2.8 below) are "equivalent". The latter describes the random periodicity in the sense of distribution. Using the theory we present in this paper, one can statistically detect periodic measures in many real world situations, with deterministic period. The deterministic period of the periodic measure can also be seen clearly in numerical experiments in [11]. To make this rigorous, in the following, we will prove directly from the pathwise description of random periodic paths, the period must be deterministic as long as the underlying noise is ergodic. This result is long over due.

To motivate a proof of the period being deterministic in the following, let us recall an observation made in [3]. Setting $\phi(s, \omega) := Y(s, \theta_{-s}\omega)$, then $Y(s + T, \omega) = Y(s, \theta_T\omega)$ for all $s \in \mathbb{R}$ if and only if $\phi(s + T, \omega) = \phi(s, \omega)$ and note also that for almost all $\omega \in \Omega$,

$$\Phi(t, \omega)\phi(s, \omega) = \phi(s + t, \theta_t\omega), \text{ for any } t, s \in \mathbb{R}, \tag{2.3}$$

is equivalent to (2.1).

Consider a random path Y of Φ. It is a function $\mathbb{R} \times \Omega \to \mathbb{X}$ satisfying $\Phi(t, \theta_s\omega)Y(s, \omega) = Y(s + t, \omega)$ for any $t \in \mathbb{R}^+$, $s \in \mathbb{R}$. At this stage we do not assume there is a constant $\epsilon > 0$ such that $Y(t + \epsilon, \omega) = Y(t, \theta_\epsilon\omega)$. Consider $\phi(s, \omega) := Y(s, \theta_{-s}\omega)$. Assume

$$T(\omega) := \inf\{t > 0 | \phi(s + t, \omega) = \phi(s, \omega) \text{ for all s }\} \tag{2.4}$$

exists. It is easy to see that for almost all $\omega \in \Omega$,

$$\phi(s + T(\omega), \omega) = \phi(s, \omega), \text{ for all } s. \tag{2.5}$$

Consider $s = T(\omega)$ in (2.3), then we have for all $\omega \in \Omega$,

$$\Phi(t, \omega)\phi(T(\omega), \omega) = \phi(t + T(\omega), \theta_t\omega) \text{ for all } t \geq 0. \tag{2.6}$$

Theorem 2.3. *Assume a measurable function $\phi : \mathbb{R} \times \Omega \to \mathbb{X}$ exists such that (2.3) holds for a.e. $\omega \in \Omega$, θ is ergodic and $T : \Omega \to \mathbb{R}^+$ defined by (2.4) exist. If T is positive $\mathbb{P}$-a.s., then it is constant $\mathbb{P}$-a.s.*

Proof. First there is a full measure set $\Omega_0 \subset \Omega$ such that (2.3), (2.5), (2.6) hold true and T is positive for all $\omega \in \Omega_0$. In the following proof we throw away a measure 0 set so when we say "all ω" we mean "all $\omega \in \Omega_0$". First from (2.3) and (2.5), for any fixed $t \geq 0$, for all ω,

$$\Phi(t,\omega)\phi(s+T(\omega),\omega) = \Phi(t,\omega)\phi(s,\omega) = \phi(s+t,\theta_t\omega) = \phi(s+t+T(\theta_t\omega),\theta_t\omega), \tag{2.7}$$

for all $s \in \mathbb{R}$. Comparing (2.6) and (2.7) we have

$$\phi(s+t+T(\theta_t\omega),\theta_t\omega) = \phi(s+t+T(\omega),\theta_t\omega) = \phi(s+t,\theta_t\omega)$$

for all $s \in \mathbb{R}$. This suggests

$$\phi(s+T(\theta_t\omega),\theta_t\omega) = \phi(s+T(\omega),\theta_t\omega) \qquad \text{for all } s \in \mathbb{R}.$$

Note for any $\omega \in \Omega$, $T(\theta_t\omega)$, by definition, should be the smallest strictly positive number satisfying

$$\phi(s+\cdot,\theta_t\omega) = \phi(s,\theta_t\omega) \qquad \text{for all } s \geq 0.$$

Thus $T(\theta_t\omega) \leq T(\omega)$.

Now we define $F_s := \{\omega : T(\omega) \leq s\}$. Then for all $t \geq 0$,

$$\theta_t^{-1}F_s = \{\omega : T(\theta_t\omega) \leq s\} \supset \{\omega : T(\omega) \leq s\} = F_s, \tag{2.8}$$

which means that F_s is a forward invariant set. Note $\theta_\cdot$ is ergodic, thus $\mathbb{P}(F_s) = 0$ or $\mathbb{P}(F_s) = 1$. Then by the definition of F_s we can conclude that T is constant $\mathbb{P}$-a.s. $\square$

In a recent paper [14], an attempt to extend T to be noise-dependent was made. The definition is quoted below. We adopt our notation for the consistency of notation in this paper.

Definition 2.4 (*Crauel Random Periodic Solution*)**.** Let $\mathbb{T} \in \{\mathbb{R}, \mathbb{R}_0^+\}$. A Crauel random periodic solution (CRPS) of random dynamical system $\Phi : \mathbb{R}^+ \times \Omega \times \mathbb{X} \to \mathbb{X}$ is a pair (ψ, T) consisting of $\mathcal{F}$-measurable functions $\psi : \Omega \times \mathbb{T} \to \mathbb{R}^m$ and $T : \Omega \to \mathbb{R}$ such that for all $\omega \in \Omega$

$$\psi(t,\omega) = \psi(t+T(\theta_{-t}\omega),\omega) \ \text{ and } \ \Phi(t,\omega)\psi(t_0,\omega) = \psi(t+t_0,\theta_t\omega)$$

for all $t, t_0 \in \mathbb{T}$.

According to Theorem 2.3, it is clear that a period that is truly random cannot exist in the sense of (2.5), which is uniform for all s. The CRPS suggestion of [14] is to allow the random period of CRPS to be different at the different part of the trajectory. That is to say that the trajectory is allowed to have returning time dependent on the starting time. This is given by some kind of random T along the pull-back path of noise in Definition 2.4.

To help understanding the difference of random periodic path and the Crauel random periodic solution, we will give three examples in which we will see that for some stochastic processes Definition 2.4 is satisfied, but these processes have no periodicity, for example, either in the pathwise sense or in the sense of distributions. Our examples show that CRPS reflects some other different kind of repeating properties, e.g. recurrence or oscillation. In particular, in the case for a deterministic function, Example 2.7 shows being CRPS is equivalent to having some oscillatory property.

Example 2.5. Consider $\Phi(t,\omega)x = x + W_t$, where $W_\cdot$ is a two-sided Brownian motion on $\mathbb{R}^1$ with $W_0 = 0$. Define $\psi(t,\omega) = -W_{-t}$, $T(\omega) = \inf\{s > 1, W_{-s} = 0\}$ and $\theta : \mathbb{R} \times \Omega \to \Omega$, $(\theta_t\omega)(s) = W_{t+s} - W_t$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. For almost all ω, $T(\omega)$ is well-defined, finite and positive, together with $T(\theta_{-t}\omega)$ for all $t \in \mathbb{R}$. First $\psi(T(\omega),\omega) = 0 = \psi(0,\omega)$. Moreover, it is easy to see that for almost all ω, $\Phi(t,\omega)\psi(t_0,\omega) = -W_{-t_0} + W_t$ and

$$\psi(t+t_0,\theta_t\omega) = -W_{-(t+t_0)}(\theta_t\omega) = -W_{-(t+t_0)+t} + W_t = -W_{-t_0} + W_t = \Phi(t,\omega)\psi(t_0,\omega),$$

and

$$\psi(t+T(\theta_{-t}\omega),\omega) = -W_{-(t+T(\theta_{-t}\omega))} = -(W_{-(t+T(\theta_{-t}\omega))} - W_{-t}) - W_{-t} = -W_{-t} = \psi(t,\omega),$$

as

$$-(W_{-(t+T(\theta_{-t}\omega))} - W_{-t}) = -W_{-T(\theta_{-t}\omega)}(\theta_{-t}\omega) = 0.$$

In conclusion, (ψ, T) satisfies Definition 2.4.

However, the random variable $T(\omega)$ ($T(\theta_{-t}\omega)$) defined in this example only reveals information about when the process $\psi(s,\omega) = -W_{-s}, s \geq 0$ (shifted process $W_{-s}(\theta_{-t}), s \geq 0$), hits zero the first time after or at time 1, i.e. some kind of recurrence. For one dimensional Brownian motion, we know such a time always exists for almost each sample path. The CRPS definition reveals that starting from any time t, after time $T(\theta_{-t}\omega)$, the process ψ will come back to the same position for the first time after time 1. But Brownian motion does not possess any periodicity and nor its law.

Example 2.6. Let us consider Brownian flow on the unit circle. Take S^1 as the group of units in $\mathbb{C}$ and the Brownian motion on the unit circle starting at 1 can be represented by $t \to e^{iW_t}$, where W is a Brownian motion on $\mathbb{R}$ with $W_0 = 0$ and the shift θ defined as in Example 2.5. As in any compact Lie group, we have a Brownian flow Φ by multiplication by such a solution [25]: $\Phi(t, \omega)z = e^{iW_t}z$, $z \in S^1$.

Define $\psi(s, \omega) := e^{-iW_{-s}}$, and define $T(\omega) := \inf\{s > 0 : W_{-s} = \pm 2\pi\}$. For almost all ω, $T(\omega)$ is well-defined, finite and positive, together with $T(\theta_{-t}\omega)$ for all $t \in \mathbb{R}$. Note

$$\Phi(t, \omega)\psi(s, \omega) = e^{iW_t - iW_{-s}} = e^{-i(W_{-(s+t)+t} - W_t)} = \psi(s + t, \theta_t\omega).$$

Moreover, by definition of $W_{-T(\omega)} = \pm 2\pi$ for almost all ω, so $W_{-T(\theta_{-s}\omega)-s} - W_{-s} = \pm 2\pi$ for almost all ω. This gives that

$$\psi(s + T(\theta_{-s}\omega), \omega) = e^{-iW_{-T(\theta_{-s}\omega)-s}} = e^{-i(W_{-s} \pm 2\pi)} = \psi(s, \omega).$$

In particular, $\psi(T(\omega), \omega) = \psi(0, \omega) = 1$.

This is a mapped Brownian motion on the unit circle and again the definition of CRPS only tells some kind of recurrence property of the process. Note Brownian motion on the unit circle has a weakly mixing invariant measure, it does not have nontrivial periodic measure.

Example 2.7. Now let us consider a deterministic case. We say a continuous real valued function $f : \mathbb{R} \to \mathbb{R}$ with $f(0) = 0$ is oscillatory around 0 to the left if there exists a real valued sequence $0 > a_1 > a_2 > \cdots$ such that $f(a_n) = 0$ for all n and $f(t) \neq 0$ for all $t \notin \{\cdots, a_n, \ldots, a_2, a_1, 0\} \cap \mathbb{R}^-$. Define $\theta_t f$, $t \in \mathbb{R}$, by $(\theta_t f)(s) = f(t+s) - f(t)$ for all $s \in \mathbb{R}$. Assume $\theta_t f$ is oscillatory around 0 to the left for all $t \in \mathbb{R}$. Define $\Phi(t, f)x = x + f(t)$, $t \geq 0$, then Φ is a (random) dynamical system as

$$\begin{aligned}
\Phi(t, \theta_s f)\Phi(s, f)x &= \Phi(t, \theta_s f)(x + f(s)) \\
&= x + f(s) + f(t + s) - f(s) \\
&= f(t + s) + x \\
&= \Phi(t + s, f)x.
\end{aligned}$$

Set $\psi(s, f) = -f(-s)$. It is easy to see that

$$\begin{aligned}
\Phi(t, f)\psi(s, f) &= -f(-s) + f(t) \\
&= -f(-(s + t) + t) + f(t) \\
&= -\left(f(-(s + t) + t) - f(t)\right) \\
&= \psi(s + t, \theta_t f).
\end{aligned}$$

Define $T(f) = \inf\{s > 0 : f(-s) = 0\}$, which is well-defined and positive due to the oscillatory assumption of f. Then $\psi(T(f), f) = -f(-T(f)) = 0 = \psi(0, f)$. Again for any t, as $\theta_{-t}f$ is oscillatory to the left around 0, so $T(\theta_{-t}f)$ exists and

$$\begin{aligned}
\psi(t + T(\theta_{-t}f), f) &= -f(-(t + T(\theta_{-t}f))) \\
&= -\left(f(-(t + T(\theta_{-t}f)) - f(-t))\right) - f(-t) \\
&= -(\theta_{-t}f)(-T(\theta_{-t}f)) - f(-t) \\
&= 0 - f(-t) \\
&= \psi(t, f).
\end{aligned}$$

So (ψ, T) satisfies Definition 2.4. But ψ is not periodic, only oscillatory.

It is noted that $T(\theta_{-t}f)$ is well defined, finite and positive for all t if and only if the function $\theta_t f$ is oscillatory around 0 to the left for all t. The "if" part has already been proved in the above. Now we prove the "only if" part. Assume $T(\theta_{-t}f)$ is well-defined for all t. Note $f(0) = 0$ and let $a_1 = 0$ and for all $n = 2, 3, \ldots$ iteratively, $a_n = -T(\theta_{-a_{n-1}}f) + a_{n-1}$ as $T(\theta_{-a_{n-1}}f)$ is well defined, finite and positive. It is easy to see from definition that f is oscillatory around 0 to the left, i.e. $f(-a_n) = 0$ for all $n = 1, 2, \ldots$. This can be done by observing $f(-a_n) = f(-a_{n-1})$ using definition of $T(\theta_{-a_{n-1}}f)$ and induction. By using a similar argument, for any t, we can construct $0 = a_1^t < a_2^t < \cdots$ such that $(\theta_t f)(-a_n^t) = 0$ for all $n = 1, 2, \ldots$.

If the function is oscillatory to the right, we can construct a backward (random) dynamical systems to obtain a similar correspondence. Both of the two cases do not imply f being periodic.

2.2. Periodic measures and law of large numbers

The main motivation of this and next subsections is to study the law of large numbers (LLN) for the subsequence random periodic path $\{Y(k\tau, \cdot)\}_{k\in\mathbb{N}}$, where τ is an arbitrary given real number that could be different from the real period T. We will obtain a number of LLNs, starting from preliminary results and eventually obtain some really nontrivial and useful results involving test periods in finding the true period, especially those in Section 2.3.

We first recall the definition of periodic measures. Define the transition probability of Markovian dynamical systems Φ as

$$P(t,x,B) := \mathbb{P}(\omega : \Phi(t,\omega)x \in B), \qquad \text{for any } B \in \mathcal{B}(\mathbb{X}).$$

Set $\mathcal{P}(\mathbb{X}) := \{\rho :$ probability measure on $(\mathbb{X}, \mathcal{B}(\mathbb{X}))\}$. The following definition was given in [3].

Definition 2.8. A measure function $\rho : \mathbb{R} \to \mathcal{P}(\mathbb{X})$ is called a periodic measure of period T on the phase space $(\mathbb{X}, \mathcal{B}(\mathbb{X}))$ for the Markovian random dynamical system Φ if it satisfies

$$\rho_{T+s} = \rho_s, \quad \text{and} \quad \rho_{s+t}(B) = \int_{\mathbb{X}} P(t,x,B)\rho_s(dx), \quad s \in \mathbb{R},\ t \in \mathbb{R}^+,\ B \in \mathcal{B}(\mathbb{X}). \tag{2.9}$$

It is called a periodic measure with minimal period T if $T > 0$ is the smallest number such that (2.9) holds. It is called an invariant measure if it also satisfies $\rho_s = \rho_0$ for all $s \in \mathbb{R}$, i.e. ρ_0 is an invariant measure for the Markovian random dynamical system Φ if

$$\rho_0(B) = \int_{\mathbb{X}} P(t,x,B)\rho_0(dx), \quad \text{for all } t \in \mathbb{R}^+,\ B \in \mathcal{B}(\mathbb{X}).$$

Let Y be a random periodic path of random dynamical system Φ, its law defined as $\rho_s(\Gamma) = \mathbb{P}\{\omega : Y(s,\omega) \in \Gamma\}$ is known to be a periodic measure [3]. The law of large numbers follows from Birkhoff's ergodic theorem under the assumption that the noise metric dynamical system is ergodic.

Assume a random periodic path Y with period $T > 0$ for the random dynamical system Φ exists. It follows from the definition of random periodic path and Birkhoff's ergodic theorem that, if $(\Omega, \mathcal{F}, \mathbb{P}, (\theta_{kT})_{k\in\mathbb{N}})$ is ergodic, then for any $\Gamma \in \mathcal{B}(\mathbb{X})$, $t \in \mathbb{R}$,

$$\frac{1}{K}\sum_{k=0}^{K-1} I_\Gamma\left(Y(t+kT,\omega)\right) = \frac{1}{K}\sum_{k=0}^{K-1} I_\Gamma\left(Y(t,\theta_{kT}\omega)\right) \to \mathbb{E}I_\Gamma(Y_t(\cdot)) = \rho_t(\Gamma) \tag{2.10}$$

as $K \to \infty$ $\mathbb{P}$-a.s. and in $L^2(\Omega, d\mathbb{P})$.

The convergence (2.10) follows from Birkhoff's ergodic theorem and the assumption that $(\Omega, \mathcal{F}, \mathbb{P}, (\theta_{kT})_{k\in\mathbb{N}})$ is ergodic immediately. However, the result itself may not be that useful in applications as the period T is often unknown and slight difference of the value T that appears on the left hand side of (2.10) can result in some significant difference to the convergence of (2.10). Thus it is crucial to study (2.10) for $Y(t+k\tau,\omega)$, where τ could be different from T. In order to study this, we will lift the metric dynamical system and we will see that the lifting enables us to prove the convergence of (2.10) even if the increment is not taken as the actual period.

Consider the metric dynamical system $(\Omega, \mathcal{F}, \mathbb{P}, (\theta_t)_{t\in\mathbb{R}})$, set $\tilde{\Omega} = [0,T) \times \Omega$, where $T > 0$ is constant and taken as the period of the random periodic path.

Note first for any fixed $t \geq 0$, there exists $m_t \in \mathbb{N}$ and $j_t \in [0,T)$ such that $t = m_tT + j_t$. For any $t \geq 0$, $(s,\omega) \in \tilde{\Omega}$, set

$$\tilde{\Theta}_t(s,\omega) = (j_{t+s}, \theta_{m_{t+s}T}\omega),$$

and for any $A \in \mathcal{B}([0,T)) \otimes \mathcal{F}$, define

$$\tilde{\mathbb{P}}(A) = \frac{1}{T}\int_{[0,T)} \mathbb{P}(A_s)ds,$$

where $A_s := \{\omega \in \Omega : (s,\omega) \in A\}$ being the s-section.

Lemma 2.9. *The map $t \mapsto \tilde{\Theta}_t$ is a semigroup and preserves $\tilde{\mathbb{P}}$.*

Proof. To prove the semigroup relation, we see that for any $t_1 \geq 0$, $t_2 \geq 0$,

$$\begin{aligned}
\tilde{\Theta}_{t_2} \circ \tilde{\Theta}_{t_1}(s,\omega) &= \tilde{\Theta}_{t_2}(j_{t_1+s}, \theta_{m_{t_1+s}T}\omega) \\
&= (j_{t_2+j_{t_1+s}}, \theta_{m_{t_2+j_{t_1+s}}T}\theta_{m_{t_1+s}T}\omega) \\
&= (j_{t_2+t_1+s}, \theta_{m_{t_2+t_1+s}T}\omega) = \tilde{\Theta}_{t_2+t_1}(s,\omega).
\end{aligned}$$

To prove the measure preserving property, note

$$\begin{aligned}
\tilde{\Theta}_t^{-1}A &= \{(s,\omega) \in \tilde{\Omega} : \tilde{\Theta}_t(s,\omega) \in A\} \\
&= \{(s,\omega) \in \tilde{\Omega} : (j_{t+s}, \theta_{m_{t+s}T}\omega) \in A\} \\
&= \{(s,\omega) \in \tilde{\Omega} : \theta_{m_{t+s}T}\omega \in A_{j_{t+s}}, s \in [0,T)\} \\
&= \{(s,\omega) \in \tilde{\Omega} : \omega \in \theta^{-1}_{m_{t+s}T}A_{j_{t+s}}, s \in [0,T)\}.
\end{aligned}$$

Then

$$\begin{aligned}\tilde{\mathbb{P}}(\tilde{\Theta}_t^{-1}A) &= \frac{1}{T}\int_{[0,T)} \mathbb{P}(\theta^{-1}_{m_{t+s}T}A_{j_{t+s}})ds \\ &= \frac{1}{T}\int_{[0,T)} \mathbb{P}(A_{j_{t+s}})ds \\ &= \frac{1}{T}\int_{[0,T)} \mathbb{P}(A_s)ds = \mathbb{P}(A). \quad \square\end{aligned}$$

This lemma enables us to apply Birkhoff's ergodic theorem.

Theorem 2.10. *Assume a random periodic path with period $T > 0$ exists. Then for any fixed s and integer $\tau > 0$, there exists a random measure function μ such that for any $\Gamma \in \mathcal{B}(\mathbb{X})$,*

$$\frac{1}{K}\sum_{k=0}^{K-1} I_\Gamma(Y(s+k\tau,\omega)) \to \mu_{s,\omega}(\Gamma), \tag{2.11}$$

$\tilde{\mathbb{P}}$-a.s. and in $L^2(\tilde{\Omega}, d\tilde{\mathbb{P}})$. Moreover, $\mu_{s,\omega} = \mu_{\tilde{\omega}} = \mu_{\tilde{\Theta}_\tau\tilde{\omega}}$.

Proof. Note

$$\begin{aligned}\frac{1}{K}\sum_{k=0}^{K-1} I_\Gamma\left(Y(s+k\tau,\omega)\right) =& \frac{1}{K}\sum_{k=0}^{K-1} I_\Gamma\left(Y(j_{s+k\tau}+m_{s+k\tau}T,\omega)\right) \\ =& \frac{1}{K}\sum_{k=0}^{K-1} I_\Gamma\left(Y(j_{s+k\tau},\theta_{m_{s+k\tau}T}\omega)\right) \\ =& \frac{1}{K}\sum_{k=0}^{K-1} I_\Gamma\left(Y(\tilde{\Theta}_{k\tau}(s,\omega))\right) \\ =& \frac{1}{K}\sum_{k=0}^{K-1} I_\Gamma\left(Y(\tilde{\Theta}_{k\tau}\tilde{\omega})\right).\end{aligned}$$

As $\tilde{\Theta}_t$ is a measurable map from $\tilde{\Omega}$ to $\tilde{\Omega}$, and preserves $\tilde{\mathbb{P}}$, so we can apply Birkhoff's ergodic theorem, there exists a random measure function μ

$$\frac{1}{K}\sum_{k=0}^{K-1} I_\Gamma\left(Y(\tilde{\Theta}_{k\tau}\tilde{\omega})\right) \to \mu_{\tilde{\omega}}(\Gamma), \qquad \text{as } k\to\infty,$$

for $\tilde{\mathbb{P}}$-a.s. $\tilde{\omega}\in\tilde{\Omega}$, and in $L^2(\tilde{\Omega}, d\tilde{\mathbb{P}})$. Thus (2.11) follows. Following again Birkhoff's ergodic theorem, it is easy to see that $\mu_{\tilde{\omega}} = \mu_{s,\omega} = \mu_{\tilde{\Theta}_\tau\tilde{\omega}}$. $\square$

Now we consider the case that τ and T are rationally dependent. Let integers q^*, p^* be co-prime to each other such that

$$q^*\tau = p^*T. \tag{2.12}$$

Then for all s,

$$s + q^*\tau = j_s + m_{s+q^*\tau}T = s + p^*T \tag{2.13}$$

and q^* is the smallest of such integer satisfying (2.13).

Following Theorem 2.10, it is easy to prove the following result.

Theorem 2.11. *Assume assumptions of Theorem 2.10 and that τ and T are rationally dependent with q^*, p^* defined by (2.12). If $\theta_{p^*T} : \Omega \to \Omega$ is ergodic, then*

$$\frac{1}{K}\sum_{k=0}^{K-1} I_\Gamma(Y(s+k\tau,\omega)) \to \mu_s(\Gamma)$$

for $\tilde{\mathbb{P}}$-a.s. $\tilde{\omega}\in\tilde{\Omega}$, and in $L^2(\tilde{\Omega}, d\tilde{\mathbb{P}})$ and μ_s is independent of ω for almost all s.

Proof. Since

$$\frac{1}{K}\sum_{k=0}^{K-1} I_\Gamma(Y(s+k\tau,\theta_{p^*T}\omega)) = \frac{1}{K}\sum_{k=0}^{K-1} I_\Gamma(Y(s+k\tau+p^*T,\omega)) = \frac{1}{K}\sum_{k=0}^{K-1} I_\Gamma(Y(s+(k+q^*)\tau,\omega)),$$

thus

$$\begin{aligned}
&\frac{1}{K}\sum_{k=0}^{K-1} I_\Gamma(Y(s+k\tau,\theta_{p^*T}\omega)) \\
&= \frac{1}{K}\sum_{k=q^*}^{K+q^*-1} I_\Gamma(Y(s+k\tau,\omega)) \\
&= \frac{1}{K}\sum_{k=0}^{K-1} I_\Gamma(Y(s+k\tau,\omega)) - \frac{1}{K}\sum_{k=0}^{q^*-1} I_\Gamma(Y(s+k\tau,\omega)) \\
&\quad + \frac{1}{K}\sum_{k=K}^{K+q^*-1} I_\Gamma(Y(s+k\tau,\omega)) \\
&\to \mu_{s,\omega}(\Gamma)
\end{aligned} \tag{2.14}$$

$\tilde{\mathbb{P}}$-a.s. and in $L^2(\tilde{\Omega}, d\tilde{\mathbb{P}})$. Here we used Theorem 2.10 in the above convergence. But $\frac{1}{K}\sum_{k=0}^{K-1} I_\Gamma(Y(s+k\tau,\theta_{p^*T}\omega)) \to \mu_{s,\theta_{p^*T}\omega}$ a.s. by Theorem 2.10 again. Thus $\mu_{s,\omega} = \mu_{s,\theta_{p^*T}\omega}$ $\tilde{\mathbb{P}}$-a.s. It follows that for almost $s \in [0,T)$, $\mu_{s,\omega} = \mu_{s,\theta_{p^*T}\omega}$ for almost all $\omega \in \Omega$. It then follows from ergodic theory as $\theta_{p^*T} : \Omega \to \Omega$ preserves $\mathbb{P}$ and is ergodic that $\mu_{s,\omega}$ is independent of ω. Thus the theorem follows immediately. □

Further analysing the rational rotation dynamical system and its periodicity, we obtain Theorem 2.12 below. For this we introduce the following notation under the assumption of Theorem 2.11, for any $s \in [0,T)$, denote

$$M_s := \{s + k\tau|T, k = 0, 1, 2, \ldots\}.$$

Then as τ and T are rationally dependent and q^*, p^* are defined in (2.13), so

$$M_s = \{s_1, s_2, \ldots, s_{q^*}\}$$

with $s \in M_s$. Let $S_l = s$ and $s_L = s + (K-1)\tau|T$ for some l and $L \in \{1, 2, \ldots, q^*\}$.

Theorem 2.12. *Assume condition of Theorem 2.11, and that τ and T are rationally dependent with q^* and p^* defined in (2.12). Then if $(\theta_{kp^*T})_{k=0,1,2,\ldots}$ is ergodic, then*

$$\frac{1}{K}\sum_{k=0}^{K-1} I_\Gamma\left(Y(s+k\tau,\omega)\right) \to \frac{1}{q^*}\sum_{i=1}^{q^*}\rho_{s_i} \qquad \text{as } K\to\infty, \quad \mathbb{P}-a.s.$$

Proof. First note that $k = q^*$ is smallest positive integer such that $s + k\tau|T = s$. Thus for each $k \le q^* - 1$, there exists unique $i_k \in \{1, 2, \ldots, q^*\}$ and $m_k \le p^* - 1$ such that

$$s + k\tau = s_{i_k} + m_k T.$$

It follows that $Y(s + k\tau, \omega) = Y(s_{i_k}, \theta_{m_k T}\omega)$.

Moreover, for each $s_i \in M_s$, there exists unique k_i and m_i such that

$$s + k_i\tau = s_i + m_i T.$$

Thus

$$\mathbb{E}\left[\sum_{k=0}^{q^*-1} I_\Gamma(Y(s+k\tau))\right] = \sum_{i=0}^{q^*-1}\rho_{s_i}(\Gamma).$$

But for any $q^* \le k \le 2q^* - 1$, note

$$s + k\tau = s + (k-q^*)\tau + q^*\tau = s + (k-q^*)\tau + p^*T,$$

it is easy to see that

$$\sum_{k=q^*}^{2q^*-1} I_\Gamma(Y(s+k\tau,\omega)) = \sum_{k=q^*}^{2q^*-1} I_\Gamma(Y(s+(K-q^*)\tau,\theta_{p^*T}\omega)) = \sum_{k=0}^{q^*-1} I_\Gamma(Y(s+k\tau,\theta_{p^*T}\omega)).$$

Similarly one can see that

$$\sum_{k=(n-1)q^*}^{nq^*-1} I_\Gamma(Y(s+k\tau,\omega)) = \sum_{k=0}^{q^*-1} I_\Gamma(Y(s+k\tau,\theta_{(n-1)p^*T}\omega)).$$

Now let N be an integer such that

$$K-1 = Nq^* + r_K, \qquad 0 \le r_K < q^*, \qquad r_K \in \mathbb{N}.$$

Then

$$\begin{aligned}
&\frac{1}{K}\sum_{k=0}^{K-1} I_\Gamma Y((s+k\tau,\omega))\\
=&\frac{1}{K}\left[\sum_{n=0}^{N-1}\sum_{k=0}^{q^*-1} I_\Gamma(Y(s+k\tau,\theta_{np^*T}\omega)) + \sum_{k=0}^{r_K} I_\Gamma(Y(s+k\tau,\theta_{Np^*T}\omega))\right]\\
=&\frac{N}{K}\sum_{k=0}^{q^*-1}\frac{1}{N}\sum_{n=0}^{N-1} I_\Gamma(Y(s+k\tau,\theta_{np^*T}\omega)) + \frac{1}{K}\sum_{k=0}^{r_K} I_\Gamma(Y(s+k\tau,\theta_{Np^*T}\omega)).
\end{aligned} \tag{2.15}$$

Note $\frac{N}{K} \to \frac{1}{q^*}$ as $K \to \infty$. So by Birkhoff's ergodic theorem, as $(\theta_{Np^*T})_{k=0,1,2,\ldots}$ is ergodic, we have as $K \to \infty$,

$$\sum_{k=0}^{q^*-1}\frac{1}{N}\sum_{n=0}^{N-1} I_\Gamma(Y(s+k\tau,\theta_{np^*T}\omega)) \to \sum_{k=0}^{q^*-1}\mathbb{E}I_\Gamma(Y(s+k\tau)) = \sum_{i=1}^{q^*}\rho_{s_i}(\Gamma). \tag{2.16}$$

Moreover, it is obvious that

$$\frac{1}{K}\sum_{k=0}^{r_K} I_\Gamma(Y(s+k\tau,\theta_{Np^*T}\omega)) \to 0. \tag{2.17}$$

Then the desired result follows from (2.15), (2.16) and (2.17) immediately. □

As a consequence, it is easy to see that

$$\mu_s = \frac{1}{q^*}\sum_{i=1}^{q^*}\rho_{s_i} \qquad \text{for any } s \in \mathbb{R}$$

is a periodic measure with period T/q^*.

In Section 3, we will consider stochastic differential equations with drift and diffusion terms depending explicitly on time t being periodic with period T. In this case, we have a stochastic semi-flow satisfying

$$u(t,r,\omega)\circ u(r,s,\omega) = u(t,s,\omega), \qquad s \le r \le t \quad \mathbb{P}-\text{a.s.}\,,$$

and random periodic condition

$$u(t+T,s+T,\omega) = u(t,s,\theta_T\omega), \qquad s \le t \quad \mathbb{P}-\text{a.s.}$$

A random path $Y : I \times \Omega \to \mathbb{X}$ is a random periodic path if $\mathbb{P}$-a.s. $u(t,s,\omega)Y(s,\omega) = Y(t,\omega)$, $s \le t$ and $Y(t+T,\omega) = Y(t,\theta_T\omega)$, $t \ge 0$. The law of large numbers that we proved for random periodic paths of random dynamical systems still holds for random periodic paths of stochastic semi-flows and the same proof works also for the stochastic semi-flow situation.

2.3. LLN of discrete time random periodic processes and arithmetic progression dynamics on a finite integer set

2.3.1. Two elementary number theory lemmas

Consider two integers $p, q \ge 1$ satisfying $p \le q$. All the results are still true when $p > q$ with a slight modification of proofs. Define $S = \{0, 1, 2, \ldots, q-1\}$, and a dynamical system on the finite integer field S, $T : S \to S$ by

$$T(i) = (i+p)|q, \qquad i \in S \tag{2.18}$$

and the trace of i as

$$S(i) = \{T^n(i) | n \in \mathbb{N}\} = \{j \in S | j = i + k_1 p - k_2 q, k_1 \in \mathbb{N}^+, k_2 \in \mathbb{N}^+ \cup \{0\}\},$$

where $i \in S$. The following two lemmas are equivalent to the linear equation Theorem on greatest common divisors (Bézout's identity)([26]). We give its equivalent result here in the language of dynamical system which is a more convenient form for the ergodic theory, the law of large numbers and the random periodic framework. This describes the dynamics on the integer field as the time index set for a time series.

Lemma 2.13. *The integers p, q are co-prime to each other if and only if $S(0) = S$.*

Proof. We consider the following proof of the result that p, q are co-prime to each other if and only if there exist integers $k_1 \geq 1, k_2 \geq 0$ such that $k_1 p - k_2 q = 1$ by studying the dynamics of the map $T : S \to S$ defined in (2.18).

First we assume that there exists $k_1 \geq 1, k_2 \geq 0$ such that $k_1 p - k_2 q = 1$. If p, q have a common divisor $r > 1$, then $r\left[k_1 \frac{p}{r} - k_2 \frac{q}{r}\right] = 1$. But this is impossible as the left hand side is a multiple of r, the right hand side is 1 only. Thus p, q must be co-prime to each other.

On the other hand, assume that p, q are co-prime to each other. We want to prove that for any $j \in S$, there exist $k_j \geq 1, l_j \geq 0$ such that $k_j p - l_j q = j$, i.e. $j \in S(0)$. If this is not the case, then statement (A): there exists $j \in S \setminus \{0\}$ such that $j \neq kp - lq$ for all $k \geq 1, l \geq 0$ holds true. Note here it is obvious that $0 \in S(0)$. Let $j_0 > 0$ be the smallest such a number. Now we want to show that $j_0 = 1$.

If $j_0 \neq 1$, then $1 \in S(0)$ and there exist $k_1 \geq 1, k_2 \geq 0$ such that $k_1 p - k_2 q = 1$. So for any $j \in S \setminus \{0\}$, $jk_1 p - jk_2 q = j$. This is a contradiction to statement (A). Thus $j_0 = 1$, i.e. $1 \notin S(0)$.

Now we consider the case that $2 \in S(0)$. In this case, it is easy to see that

$$\{2, 4, \ldots\} \cap S \subset S(0). \tag{2.19}$$

However, let $k_2 \geq 1, l_2 \geq 0$ be such numbers that $k_2 p - l_2 q = 2$, then $1 + k_2 p - l_2 q = 3$. Thus $3 \in S(1)$. Generally, we can easily see that

$$\{3, 5, \ldots\} \cap S \subset S(1). \tag{2.20}$$

From (2.19), (2.20) and the fact that $1 \notin S(0)$ we know that

$$S(0) = \{0, 2, 4, \ldots\} \cap S, \quad \text{and} \quad S(1) = \{1, 3, 5, \ldots\} \cap S.$$

But it is obvious that $p \in S(0)$. So p must be an even number. We claim that q is even as well. For this, let k be the smallest integer such that $kp > q$. Then $kp - q \in S(0)$. So $kp - q$ is an even number. Since kp is even, so q must be even. Thus p, q have a common divisor 2.

Next we consider the case that $2 \notin S(0)$, but $3 \in S(0)$. Similar to the above argument, we know that

$$S(0) = \{0, 3, 6, \ldots\} \cap S, \quad S(1) = \{1, 4, 7, \ldots\} \cap S, \quad \text{and} \quad S(2) = \{2, 5, 8, \ldots\} \cap S.$$

Clearly, $p \in S(0)$, so p is divisible by 3.

Similarly as above, let k be the smallest integer such that $kp > q$. Then $kp - q \in S(0)$. Since p, $kp - q$ are divisible by 3, q must be divisible by 3 as well. This is the case we conclude p, q have common divisor 3.

In general, we can prove that if $r > 1$ is an integer such that $2, 3, \ldots, r - 1 \notin S(0)$ but $r \in S(0)$, then p, q have a common divisor r.

Summarising above, we can say that if $1 \notin S(0)$, then p, q have a common divisor greater than 1. This contradicts with the assumption that p, q are co-prime. That means $1 \in S(0)$ and $S(0) = \{0, 1, 2, \ldots\} \cap S = S$ is proved. □

From Lemma 2.13 and its proof, it is easy to see that p, q are co-prime to each other if and only if $S(0) = S(i) = S = \{0, 1, 2, \ldots, q - 1\}$ for all $i \in S$. But when p, q are not co-prime to each other, we have the following result.

Lemma 2.14. *The integers p, q have a greatest common divisor r if and only if for $0 \leq i < r$,*

$$S(i) = \{i, r + i, 2r + i, \ldots\} \cap S.$$

Proof. "Only if" part: Assume p, q have a greatest common divisor r. When $r = 1$, the result is Lemma 2.13. When $r > 1$, there exist co-prime integers p^*, q^* such that $p = rp^*, q = rq^*$. But from Lemma 2.13, we know that for any $i^* \in \{0, 1, 2, \ldots, q^* - 1\} =: S^*$, there exist $k^* \geq 1, l^* \geq 0$ such that $k^* p^* - l^* q^* = i^*$. So $k^* p - l^* q = ri^*$. This means $ri^* \in S(0)$. In particular, $r \in S(0)$.

Next we want to show that for any integer $0 < s < r$, $s \notin S(0)$. If it is not the case, say $s \in S(0)$. This also says that there exist $k \geq 1$ and $l \geq 0$ such that $kp - lq = s$. So $r\left[k\frac{p}{r} - l\frac{q}{r}\right] = s$. This is a contradiction as the left hand side is a multiple of r, while the right hand side is not. This suggests that $S(0) = \{0, r, 2r, \ldots\} \cap S$. The general result follows easily.

The "if" part was directly contained in the proof of Lemma 2.13. □

Remark 2.15. This result is the integer field version of the dynamical system of rotation on unit circle S^1 by an angle γ. In the classical dynamical system theory, the rotation system is ergodic on the unit circle with the Lebesgue measure when $\frac{\gamma}{2\pi}$ is irrational, and is periodic (non-ergodic) when $\frac{\gamma}{2\pi}$ is rational. In the case considered in this paper, the dynamical system defined by (2.18) is ergodic on S with the uniform distribution if and only if p, q are co-prime to each other. Otherwise S is divisible into distinct ergodic components $S(0) = \{0, r, 2r, \ldots\} \cap S$, $S(1) = \{1, r + 1, 2r + 1, \ldots\} \cap S, \cdots$, $S(r - 1) = \{r - 1, 2r - 1, \ldots\} \cap S$. It is easy to know $T : S(i) \to S(i)$ is ergodic for each i. We will explain soon that this number theory result which is presented in Lemmas 2.13 and 2.14 as a dynamical system result has significance in finding the period of random periodic processes by a time series using the law of large numbers.

2.3.2. Law of large numbers through arithmetic progression of test period

Now we consider discrete time random periodic process $Y : I \times \Omega \to \mathbb{R}$, where $I = \{\ldots, -1, 0, 1, \ldots\}$, and denote its law by $\rho_s \in \mathcal{P}(\mathcal{B}(\mathbb{R}))$, $s \in I$,

$$\rho_s(A) = P\{\omega : Y(s, \omega) \in A\}, \qquad A \in \mathcal{B}(\mathbb{R}),$$

and assume there exists $q \in \mathbb{N}$ such that $\rho_{s+q} = \rho_s$, $s \in \mathbb{Z}$.

For convenience, we extend $S(i)$, $i \in S$ to $i \in \mathbb{N}$ by setting $S(i) = S(i|q)$. Then

$$S(0) = S(q) = S(2q) = \cdots, \quad \text{and} \quad S(1) = S(1+q) = S(1+2q) = \cdots.$$

We obtain the following law of large numbers along arithmetic progression with an arbitrary common increment as a test period. For integer $p > 0$, regarded as the test period, denote by r the common divisor of p and the true period q and $p^* = \frac{p}{r}$, $q^* = \frac{q}{r}$. Then $S(i) = \{i, i+r, i+2r, \ldots, i+(q^*-1)r\}$, $i = 0, 1, \ldots, r-1$; $S(i+1) = S(i)$, $i = 0, 1, \ldots$.

Theorem 2.16. *Let Y be a random periodic path of the discrete time random dynamical system Φ (or a periodic stochastic semi-flow u), with the period $T = q$. Assume the metric dynamical system $(\Omega, \mathcal{F}, \mathbb{P}, (\theta_{kp^*q})_{k\in\mathbb{N}})$ is ergodic. Then for any $i \in \mathbb{N}$, $A \in \mathcal{B}(\mathbb{R})$, $p \in \mathbb{N}^+$,*

$$\frac{1}{K}\sum_{k=0}^{K-1} I_A(Y_{i+kp}(\omega)) \to \frac{1}{q^*}\sum_{u\in S(i)} \rho_u(A) \quad \text{as } K \to \infty,$$

almost surely and in $L^2(\Omega, d\mathbb{P})$.

In particular, when p, q are co-prime,

$$\frac{1}{K}\sum_{k=0}^{K-1} I_A(Y_{i+kp}(\omega)) \to \frac{1}{q}\sum_{u=0}^{q-1} \rho_u(A) \quad \text{as } K \to \infty$$

and when $p = q$,

$$\frac{1}{K}\sum_{k=0}^{K-1} I_A(Y_{i+kp}(\omega)) \to \rho_i(A) \quad \text{as } K \to \infty,$$

almost surely and in $L^2(\Omega, d\mathbb{P})$.

Proof. Consider $i = 0, 1, 2, \ldots, q-1$ first. Set $\xi_k^i = Y_{i+kp}$, $i = 0, 1, \ldots, r-1$ and $i + kp = mq + u$, $0 \le u < q$, $u \in S(i)$, then

$$Y_{i+kp}(\omega) = Y_u(\theta_{mq}\omega), \qquad u \in S(i),$$

and

$$\sum_{k=0}^{q^*-1} \rho_{i+kp}(A) = \sum_{u=i,i+r,\ldots,i+(q^*-1)r} \rho_u(A). \tag{2.21}$$

It is also important to note that for any $u \in S(i)$, there exists a unique $0 \le k \le q^*-1$ such that

$$i + kp = m_{i+kp,u}q + u.$$

Thus $m = m_{i+kq,u}$ is also unique for each $u \in S(i)$.

Now note $q^*p = p^*q$, so

$$\sum_{k=q^*}^{2q^*-1} I_A(Y_{i+kp}(\omega)) = \sum_{k=0}^{q^*-1} I_A(Y_{i+kp+p^*q}(\omega)) = \sum_{k=0}^{q^*-1} I_A(Y_{i+kp}(\theta_{p^*q}\omega)).$$

Similarly, for general $n \ge 1$,

$$\sum_{k=(n-1)q^*}^{nq^*-1} I_A(Y_{i+kp}(\omega)) = \sum_{k=0}^{q^*-1} I_A(Y_{i+kp}(\theta_{(n-1)p^*q}\omega)).$$

Thus

$$\frac{1}{Nq^*}\sum_{k=0}^{Nq^*-1} I_A(Y_{i+kp}(\omega)) = \frac{1}{N}\sum_{n=0}^{N-1}\frac{1}{q^*}\sum_{k=0}^{q^*-1} I_A(Y_{i+kp}(\theta_{(n-1)p^*q}\omega))$$
$$=\frac{1}{q^*}\sum_{k=0}^{q^*-1}\frac{1}{N}\sum_{n=1}^{N} I_A(Y_{i+kp}(\theta_{(n-1)p^*q}\omega)).$$

For fixed i and k, as $(\theta_{(n-1)p^*q})_{n=1,2,\ldots}$ is ergodic, so by Birkhoff's ergodic theorem,

$$\frac{1}{q^*}\sum_{k=0}^{q^*-1}\frac{1}{N}\sum_{n=1}^{N} I_A(Y_{i+kp+(n-1)p^*q}(\omega)) \xrightarrow{N\to\infty} \frac{1}{q^*}\sum_{k=0}^{q^*-1}\rho_{i+kp}(A) \quad = \frac{1}{q^*}\sum_{u\in S(i)}\rho_u(A) \tag{2.22}$$

a.s. and in $L^2(\Omega, d\mathbb{P})$. Note here $S(i) = \{i, i+r, \ldots, i+(q^*-1)r\}$, $i = 0, 1, \ldots, r-1$, and we used (2.21). The proof of the theorem follows from the above and the extension of $S(i)$ to $i \in \mathbb{N}$.

The result for periodic stochastic semi-flows can be proved by exactly the same argument. □

Remark 2.17. It is noted that for a stochastic system with random periodic path, if the underlying noise and the associated measure preserving dynamical system is ergodic, then the law of large numbers in Theorem 2.16 holds. In many situations, (e.g. Examples 3.1 and 3.3) the underlying noise is Brownian motion. In this case the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a Wiener space and the measure preserving dynamical system $\theta : I \times \Omega \to \Omega$ is given by $(\theta_t\omega)(s) = W(t+s) - W(t)$.

In [27], it was proved that the metric dynamical system given as the shift of Brownian motions is ergodic. The theorem is stated below.

Theorem 2.18. *The canonical dynamical system of Brownian motion $\Sigma = (\Omega, \mathcal{F}, \mathbb{P}, (\theta_t)_{t\in\mathbb{T}})$ ($\mathbb{T} = \mathbb{R}^+$ or $\mathbb{R}$) and their discrete dynamical system $\Sigma^T = (\Omega, \mathcal{F}, \mathbb{P}, (\theta_T^n)_{n\in I})$ ($I = \mathbb{N}$ or $\mathbb{Z}$) are ergodic.*

Theorem 2.18 suggests that the assumption in Theorem 2.16 can be guaranteed by testing the distribution of noise being a Gaussian white noise. We will show that for a given dataset in Example 3.4 and Remark 3.5 by some numerical method in the next section.

2.4. Convergence of time average of empirical measures in Wasserstein distance

In the following we will establish a scheme to detect the true period q of dataset $\{y_t\}_{t=1}^n$ by comparing laws of $\{y_{i+kp}\}_{k=0}^{N-1}$ with a test period p being unnecessarily equal to q. Details of the EPM scheme and numerical experiments will be shown in Section 3. Numerically it is convenient to use the Wasserstein distance to quantify the periodicity of the laws of random periodic process. As we can only use the empirical measures in the numerical scheme, next we will prove the convergence of the empirical measures to the true measure under the Wasserstein metric. For this, first we introduce some notation. Let $d \geq 1$ and $\mathcal{P}(\mathbb{R}^d)$ be the set of all probability measures on $\mathbb{R}^d$. For $\alpha \geq 1$ and $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$, consider the αth Wasserstein distance between them as

$$W_\alpha(\mu, \nu) := \inf_{\xi\in\mathcal{H}(\mu,\nu)}\left\{\left(\int_{\mathbb{R}^d\times\mathbb{R}^d}|x-y|^\alpha\xi(dx, dy)\right)^{\frac{1}{\alpha}}\right\},$$

where $\mathcal{H}(\mu, \nu)$ is the set of all probability measures on $\mathbb{R}^d \times \mathbb{R}^d$ with marginals μ and ν. For $0 < \alpha < 1$,

$$W_\alpha(\mu, \nu) := \inf_{\xi\in\mathcal{H}(\mu,\nu)}\left\{\left(\int_{\mathbb{R}^d\times\mathbb{R}^d}|x-y|^\alpha\xi(dx, dy)\right)\right\}.$$

It is very natural to use the Wasserstein distance to describe the periodicity of a periodic measure and to detect periodicity in a time series dataset, as $i \mapsto W_\alpha(\mu_1, \mu_i)$ is a real valued periodic function. In this subsection, we will establish the theoretical result on the convergence of empirical distributions in the Wasserstein distance. In Section 3, we will use this result to deal with data.

In order to prove the main result of this section (Theorem 2.20), we recall the following result in [28].

Consider first on $(-1, 1]^d$, denote $\mathcal{P}_l$ as the natural partition of $(-1, 1]^d$ into 2^{dl} disjoint, equal-distance sets. For example, when $d = 1$, $\mathcal{P}_l = \{(-1+\frac{k}{2^{l-1}}, -1+\frac{k+1}{2^{l-1}}]\}_{k=0}^{2^l-1}$. To extend to $\mathbb{R}^d \times \mathbb{R}^d$, we introduce $B_0 := (-1, 1]^d$ and $B_n := (-2^n, 2^n]^d \backslash (-2^{n-1}, 2^{n-1}]^d$ for $n \geq 1$. In [28], the authors proved the following lemma:

Lemma 2.19. *Let $d \geq 1$ and $\alpha > 0$. For all pairs of probability measures μ, ν on $\mathbb{R}^d$,*

$$W_\alpha^\alpha(\mu, \nu) \leq \kappa_{\alpha,d}C_\alpha\sum_{n\geq0}2^{\alpha n}\sum_{l\geq0}2^{-\alpha l}\sum_{F\in\mathcal{P}_l}|\mu(2^nF\cap B_n) - \nu(2^nF\cap B_n)|, \tag{2.23}$$

with the notation $2^nF := \{2^nx : x \in F\}$ and where $\kappa_{\alpha,d} := 2^\alpha d^{\alpha/2}(2^\alpha+1)/(2^\alpha-1)$ and $C_\alpha := 1 + 2^{-\alpha}/(1-2^{-\alpha})$.

For fixed i, p and $A \subset \Omega$, set

$$\mu_K^{i,p} := \frac{1}{K}\sum_{k=0}^{K-1}\delta_{Y_{i+kp}}$$

as the empirical measure generated by the random periodic path Y, and $\mu^{i,p} := \frac{1}{q^*}\sum_{u\in S(i)}\rho_u$. In particular, Lemma 2.19 gives

$$W_1(\mu_K^{i,p},\mu^{i,p}) \le \kappa_{1,d}C_1\sum_{n\ge 0}2^n\sum_{l\ge 0}2^{-l}\sum_{F\in\mathcal{P}_l}|\mu_K^{i,p}(2^nF\cap B_n)-\mu^{i,p}(2^nF\cap B_n)|. \tag{2.24}$$

We will prove the following main result of this subsection.

Theorem 2.20. *Assume all the conditions of Theorem 2.16 hold and there exists $\delta > 0$ such that for all t, $\int_{\mathbb{R}}|x|^{\delta+1}\rho_t(dx) < \infty$. Then as $K \to \infty$,*

$$\mathbb{E}[W_1(\mu_K^{i,p},\mu^{i,p})] \to 0.$$

Proof. As there exists $\delta > 0$ such that $\int_{\mathbb{R}}|x|^{\delta+1}\rho_t(x) < \infty$, it is obvious that for any i,p, $M_{\delta+1}(\mu^{i,p}) := \int_{\mathbb{R}}|x|^{\delta+1}\mu^{i,p}(x) < \infty$. Then by Chebyshev inequality, for all $n \ge 0$, $\mu^{i,p}(B_n) \le P_{\mu^{i,p}}(|X| > 2^{n-1}) \le M_{\delta+1}(\mu^{i,p})2^{-(\delta+1)(n-1)}$.

Consider the expectation of (2.24),

$$\mathbb{E}W_1(\mu_K^{i,p},\mu^{i,p}) \le \kappa_{1,d}C_1\sum_{n\ge 0}2^n\sum_{l\ge 0}2^{-l}\sum_{F\in\mathcal{P}_l}\mathbb{E}|\mu_K^{i,p}(2^nF\cap B_n)-\mu^{i,p}(2^nF\cap B_n)|. \tag{2.25}$$

Denote $f_{n,l,K} := \sum_{F\in\mathcal{P}_l}\mathbb{E}|\mu_K^{i,p}(2^nF\cap B_n)-\mu^{i,p}(2^nF\cap B_n)|$. Suppose for each $K > q^*$, we can decompose it as $K = j + k'q^*$, where $j \in \{0,1,\ldots,q^*-1\}$ and $k' \in \mathbb{Z}^+$. Then for $K > q^*$,

$$\begin{aligned}
f_{n,l,K} &\le \sum_{F\in\mathcal{P}_l}\left(\frac{1}{K}\sum_{k=0}^{K-1}\mathbb{E}\delta_{Y_{i+kp}}(2^nF\cap B_n)+\mu^{i,p}(2^nF\cap B_n)\right)\\
&\le \sum_{F\in\mathcal{P}_l}\left(\mu^{i,p}(2^nF\cap B_n)+\frac{q^*}{K}\mu^{i,p}(2^nF\cap B_n)+\mu^{i,p}(2^nF\cap B_n)\right)\\
&\le 3\sum_{F\in\mathcal{P}_l}\mu^{i,p}(2^nF\cap B_n)\\
&\le 3\mu^{i,p}(B_n) \le 3M_{\delta+1}(\mu^{i,p})2^{-(\delta+1)(n-1)},
\end{aligned}$$

and

$$\begin{aligned}
&\sum_{n\ge 0}2^n\sum_{l\ge 0}2^{-l}\left(3M_{\delta+1}(\mu^{i,p})2^{-(\delta+1)(n-1)}\right)\\
&= 6M_{\delta+1}(\mu^{i,p})2^{\delta+1}\sum_{n\ge 0}2^n2^{-(\delta+1)n} = 6M_{\delta+1}(\mu^{i,p})\frac{2^{\delta+1}}{1-2^{-\delta}} < \infty
\end{aligned}$$

Also, by the Cauchy–Schwartz inequality and Theorem 2.16 we have that for any fixed n, l,

$$0 \le f_{n,l,K} \le \sum_{F\in\mathcal{P}_l}\left(\mathbb{E}|\mu_K^{i,p}(2^nF\cap B_n)-\mu^{i,p}(2^nF\cap B_n)|^2\right)^{\frac{1}{2}} \to 0,\ \text{as } K\to\infty.$$

Hence by the Dominated Convergence Theorem, we obtain that as $K \to \infty$,

$$0 \le \mathbb{E}[W_1(\mu_K^{i,p},\mu^{i,p})] \le \kappa_{1,d}C_1\sum_{n\ge 0}2^n\sum_{l\ge 0}2^{-l}f_{n,l,K} \to 0.$$

So the Theorem follows. □

The following corollaries show that in the numerical scheme, we can use $W_1(\mu_K^{i,p},\mu_K^{i+h,p})$ as the approximation of $W_1(\mu^{i,p},\mu^{i+h,p})$.

Corollary 2.21. *Assume all the conditions of Theorem 2.20 hold. There exists a subsequence $K_m \to \infty$ as $m \to \infty$ such that $W_1(\mu_{K_m}^{i,p},\mu_{K_m}^{j,p}) \to W_1(\mu^{i,p},\mu^{j,p})$ as $m\to\infty$ for all $i,j \in \{0,1,\ldots,r-1\}$ a.s.*

Proof. For any fixed $i, j \in \{0, 1, \ldots, r-1\}$, by the triangle inequality of Wasserstein distance, we have

$$W_1(\mu_K^{i,p}, \mu_K^{j,p}) \le W_1(\mu_K^{i,p}, \mu^{i,p}) + W_1(\mu^{i,p}, \mu^{j,p}) + W_1(\mu^{j,p}, \mu_K^{j,p}), \tag{2.26}$$

and

$$W_1(\mu^{i,p}, \mu^{j,p}) \le W_1(\mu_K^{i,p}, \mu^{i,p}) + W_1(\mu_K^{i,p}, \mu_K^{j,p}) + W_1(\mu_K^{j,p}, \mu^{j,p}). \tag{2.27}$$

Note that, from Theorem 2.20, we have that $\mathbb{E}[W_1(\mu_K^{i,p}, \mu^{i,p})] \to 0$ and $\mathbb{E}[W_1(\mu_K^{j,p}, \mu^{j,p})] \to 0$ as $K \to \infty$, so by a classical Chebyshev inequality and Borel–Cantelli Lemma argument, there exists a subsequence $K_m \to \infty$ as $m \to \infty$ such that $W_1(\mu_{K_m}^{i,p}, \mu^{i,p}) \to 0$ and $W_1(\mu_{K_m}^{j,p}, \mu^{j,p}) \to 0$ as $m \to \infty$ a.s.

Thus it follows from (2.26) and (2.27) that as $m \to \infty$,

$$W_1(\mu_{K_m}^{i,p}, \mu_{K_m}^{j,p}) \to W_1(\mu^{i,p}, \mu^{j,p}) \quad \text{a.s.} \tag{2.28}$$

Finally, it is trivial to extend (2.28) for any fixed $i, j \in \{0, 1, \ldots, r-1\}$ to all $i, j \in \{0, 1, \ldots, r-1\}$. $\square$

Corollary 2.22. *Assume all conditions of Theorem 2.20 hold. If $W_1(\mu_K^{i,p}, \mu_K^{j,p})$ has a limit as $K \to \infty$ a.s., then*

$$\lim_{K\to\infty} W_1(\mu_K^{i,p}, \mu_K^{j,p}) = W_1(\mu^{i,p}, \mu^{j,p}) \quad \textit{a.s.}$$

3. Detecting periodicity and dataset experiments

3.1. Periodic measures and Wasserstein distance versus DFT

A commonly used method to estimate periodicity of a dataset is periodogram. See relevant statements in [29] and our comments in the first paragraph of the Introduction. For a dataset $\{y(t)\}_{t=0}^{n-1}$, compute its discrete Fourier transformation (DFT) as $d(f) = \frac{1}{n}\sum_{t=0}^{n-1} y(t)e^{-2\pi i f t} = \frac{A(f)}{2} - i\frac{B(f)}{2}$, where $A(f) = \frac{2}{n}\sum_{t=0}^{n-1} y(t)\cos(2\pi f t)$ and $B(f) = \frac{2}{n}\sum_{t=0}^{n-1} y(t)\sin(2\pi f t)$, and f takes values from $0, \frac{1}{n}, \frac{2}{n}, \ldots, \frac{n}{2n}$ (the only frequencies that need to be considered due to aliasing effect). The periodogram of frequency f is defined as $I(f) = n|d(f)|^2$. It measures how strongly the oscillation with frequency f is represented in the data. Then the dominated period of the dataset is estimated as $1/f_{\max}$ where $I(f)$ reaches its maximum at $f_{\max}$. Fisher's test and Fuller's test give the significance of the largest peak in a periodogram plot to guide whether or not to accept the estimate period. The idea of this method is to use Fourier transformation (or wavelets transformation) as a filter to test which frequency component of the filter will interact greatest with the dataset. It is actually to approximate the dataset by using deterministic periodic functions.

In this section, we will construct a new scheme of detecting periodicity from concerning the periodic measures of random periodic processes. We will see in Examples 3.1 and 3.3 that the classical DFT method may fail to detect the actual period robustly in some cases especially when the periodicity is weak relative to the noise or when the noise may have periodicity with different period from that of the trend process. We will demonstrate that the ergodic periodic measure (EPM) scheme we propose here can work efficiently in these situations, thus provides some new insight to the analysis of time series. This can be useful potentially in many applications.

Consider an integer $p \in \mathbb{N}$ and assume the period of the random periodic process is $q \in \mathbb{N}$. Recall Theorems 2.16, 2.20 and their statistical implication to consider empirical measure $\mu_K^{i,p}$ to approximate $\mu^{i,p} = \frac{1}{q^*}\sum_{u\in S(i)} \rho_u(A)$, where q^* is the integer such that $q = rq^*$ with $r = gcd(p, q)$. Assume at the moment that q is known. We learnt from Theorem 2.16 that as the map $i \mapsto S(i)$ is periodic with period r, so the function $i \mapsto \mu^{i,p}$ is periodic with period r. In particular, when p, q are co-prime, $i \mapsto \mu^{i,p}$ is aperiodic and $\mu^{i,p}$'s are equal for all i. In contrast, when $p = q$, $i \mapsto \mu^{i,p}$ is periodic with period q.

We use Wasserstein distance "*wasserstein1d*$(\cdot)$" in the package *transport* in the R language to quantitatively measuring the difference between two measures in one dimension. The approximation of the empirical measure to the periodic measure quantified in the sense of Wasserstein distance was studied in Theorem 2.20, Corollaries 2.21 and 2.22. The Wasserstein distance is a convenient way numerically to measure the difference or the evaluation of $t \mapsto \rho_t$. The Wasserstein distance between two datasets $\{x_i\}_{i=1}^m$ and $\{y_j\}_{j=1}^n$ is computed as follows: first we approximate the empirical distribution functions $F^{(x)}(t) = \frac{1}{m}\sum_{i=1}^m I_{[x_i,\infty)}(t)$ for dataset $\{x_i\}_{i=1}^m$ and $G^{(y)}(t) = \frac{1}{n}\sum_{j=1}^n I_{[y_i,\infty)}(t)$ for dataset $\{y_j\}_{j=1}^n$, then the αth-Wasserstein distance is given as $W_\alpha(F, G) = \left(\int_0^1 |F^{-1}(u) - G^{-1}(u)|^\alpha du\right)^{1/\alpha}$, where F^{-1} and G^{-1} are generalised inverses. Specially when $m = n$, $W_\alpha(F, G) = \left(\frac{1}{n}\sum_{i=1}^n |x_i - y_i|^\alpha\right)^{\frac{1}{\alpha}}$, where both $\{x_i\}_{i=1}^m$ and $\{y_j\}_{j=1}^n$ are reordered in the ascending order. See [30] for statistical aspects of the Wasserstein distance.

The more different between $F^{(x)}$ and $G^{(y)}$, the bigger $W_1(F, G)$ is, and $W_1(F, G) = 0$ if $F^{(x)}$ is the same with $G^{(y)}$. Note here in the computations of Wasserstein distance, we normalise both the datasets $\{x_i\}_{i=1}^n$ and $\{y_i\}_{i=1}^n$ by dividing the total mean of the two datasets $\frac{1}{2n}\sum_{i=1}^n (x_i + y_i)$ in order to control the overall bound of the Wasserstein distance. This makes sense in this computation as their relative values are more important and the normalisation does not change their order.

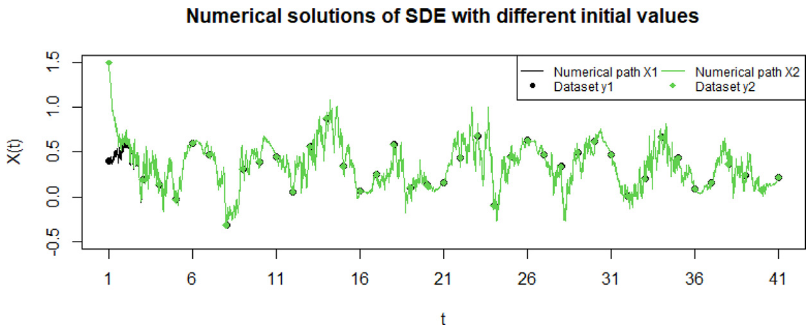


Fig. 3.1. Numerical simulation of the solution to SDE (3.1).

We use the following example to compare the classical DFT and our new periodic measure approach.

Example 3.1. Consider the following stochastic differential equation (SDE),

$$dX(t) = \left(-\pi X(t) + \sin\left(\frac{\pi t}{2}\right) + 1\right) dt + \left(0.1 + 0.6 \sin\left(\frac{\pi t}{5}\right)\right) dW_t, \tag{3.1}$$

with $X(0) = x$, where $W_t, t \in \mathbb{R}$, is a two-sided Brownian motion on $\mathbb{R}^1$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The solution is denoted by $X(t, x)$.

(a) **Explicit periodic measures**. From [4], we know that the infinite horizon process

$$Y(t) = \int_{-\infty}^{t} f(s)e^{-\pi(t-s)}ds + \int_{-\infty}^{t} g(s)e^{-\pi(t-s)}dW_s, \tag{3.2}$$

where $f(s) := \sin\left(\frac{\pi s}{2}\right) + 1$ and $g(s) := 0.1 + 0.6 \sin\left(\frac{\pi t}{5}\right)$, is a random periodic solution of SDE (3.1) with period $T = 20$. For any initial value x, $|X(t + kT, 0, \theta_{-kT}\omega, x) - Y(t, \omega)| \to 0$ as $k \to \infty$ a.s. Here $\theta : \mathbb{R} \times \Omega \to \Omega$ is given by $(\theta_t\omega)(s) = W_{t+s}(\omega) - W_t(\omega)$, for any $s, t \in \mathbb{R}$. The pull-back convergence is very convenient when the pathwise convergence is considered [31]. From the analysis of Example 4.9 in [9], we obtain by a similar calculation that the periodic measure ρ_t of SDE (3.1) exists and is a Gaussian distribution with mean $\frac{1}{\pi} + 0.8\left(\frac{\sin\left(\frac{\pi t}{2}\right)}{\pi} - \frac{\cos\left(\frac{\pi t}{2}\right)}{2\pi}\right)$ and variance $\frac{0.1^2}{2\pi} + (2 \times 0.1 \times 0.6) \times \frac{100}{101}\left(\frac{\sin\left(\frac{\pi t}{5}\right)}{2\pi} - \frac{\cos\left(\frac{\pi t}{5}\right)}{20\pi}\right) + \frac{0.6^2}{2}\left(\frac{1}{2\pi} - \frac{25}{26}\left(\frac{\cos\left(\frac{2\pi t}{5}\right)}{2\pi} + \frac{\sin\left(\frac{2\pi t}{5}\right)}{10\pi}\right)\right)$.

We can see that the periodicities of the mean with period 4 and of the noise fluctuation with period 10 contribute to the periodicity of the whole process. Hence the real period of the process is 20.

(b) **Data generation**. We use the numerical method introduced in [32] to simulate the random periodic path of SDE (3.1), thus to generate a synthetic dataset. In fact, we can consider SDE (3.1) starting at any time $s \in \mathbb{R}$, the solution is denoted by $X(t, s, \omega, x)$. Note $X(t + kT, 0, \theta_{-kT}\omega, x) = X(t, -kT, \omega, x)$ for any $k \in \mathbb{N}$ a.s. Thus we consider $X_t^{-kT} := X(t, -kT, \omega, x)$ without involving θ in the numerical computation. The starting time of the approximating scheme is taken to be $-kT$ for some fixed positive integer k. In fact we only need to consider $-kT \leq t \leq 0$. The time domain from $-kT$ to 0 is divided into N intervals of length Δt such that $N\Delta t = kT$. Then the iteration formula is

$$\begin{aligned}\hat{X}_{-kT+(i+1)\Delta t}^{-kT} =& \hat{X}_{-kT+i\Delta t}^{-kT} + f(i\Delta t, \hat{X}_{-kT+i\Delta t}^{-kT})\Delta t \\ &+ g(i\Delta t, \hat{X}_{-kT+i\Delta t}^{-kT})(W_{-kT+(i+1)\Delta t} - W_{-kT+i\Delta t}),\end{aligned} \tag{3.3}$$

where $i = 0, 1, \ldots, N-1$ and $\hat{X}_{-kT+0\Delta t}^{-kT} = x$. It was proved in [32] that the numerical solution converges strongly as $k \to \infty$ to the random periodic path of the approximated system. The convergence to the approximate random periodic path is fairly fast in a speed of exponential convergence in this case. This is an approximation to the exact random periodic path with an error in the order of $(\Delta t)^{\frac{1}{2}}$ in the L^2-norm. In the numerical scheme, we choose the step size Δt to be 0.01, which is small enough to ensure the convergence of empirical measures to theoretic periodic measures with an error in the order of Δt in the weak topology by the analysis in [11]. So there are 2000 points within each period. We produce two paths with different initial values under the same noise path to verify the convergence, and part of them is plotted in Fig. 3.1. The black curve represents $\hat{X}_1(t)$ with initial value $\hat{X}_1(0) = 0.4$ and the green one represents $\hat{X}_2(t)$ with initial value $\hat{X}_2(0) = 1.5$. We can see from Fig. 3.1 that the two paths merge together quickly.

Thus we can use the simulation path $\hat{X}_1$ to test our EPM scheme. The length of the dataset is $N^* = 100,000$ periods long to ensure that the estimate of empirical density functions in the following analysis will be accurate enough. We only select 20 equal-time-interval points in each period to form a discrete dataset $\{y(t)\}$ of length 2 million which has period 20 for simplicity. The period 20 of the discrete dataset $\{y(t)\}$ is the same as in the real time case. The black spots in Fig. 3.1

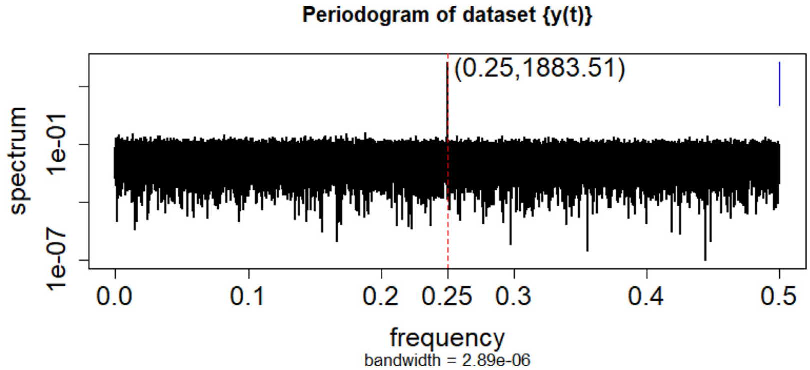


Fig. 3.2. Periodogram of dataset generated from SDE (3.1).

represent the sub-dataset of the first 2 periods. This selected dataset $\{y(t)\}$ of discrete time has relatively small period that makes our experiment easier to handle. It gives a typical example of time series of random periodic process.

(c) **The shortfall of DFT approach**. In this case the real period of the data is $q = 20$. We first apply DFT to detect q. We use the function "*spec.pgram*$(\cdot)$" in the package *STATS* in R Language in which it calculates the periodogram using a fast Fourier transform, and optionally smooths the result with a series of modified Daniell smoothers (moving averages giving half weight to the end values). The plot of periodogram is shown in Fig. 3.2. The maximum spectrum, which is 1883.51, happens at frequency 0.25, hence the estimated dominated period of this dataset is $1/0.25 = 4$, but the real period of the dataset is 20. In fact, from [4,9] and our analysis in the beginning of this example, the SDE (3.1) has a random periodic path and a periodic measure of period 20. The periodicity reflecting to the discrete dataset is 4 on the mean trend and 10 on the noise fluctuations respectively. This suggests that the DFT method only detects the periodicity of the mean trend but not sensitive to the periodicity of the noise.

In such a case the DFT method fails to detect the true period of the dataset even with reasonable amount of data (2,000,000), but the EPM scheme can reveal the clear periodicity of period 20. It surprised us when we first observed that phenomenon. Now we can well understand it as the idea of DFT is to use a deterministic periodic Fourier series to approximate a time series in a pathwise sense, while the idea of EPM is to detect the evolution of the distribution of the time series. The assumption of DFT is the times series can be decomposed numerically as $y(t) \approx f(t) + z(t)$, where $f(t)$ is a deterministic periodic function and $z(t)$ is a stationary process. This assumption is not satisfied in many situations. The main purpose of DFT is to estimate the function $f(t)$ but it cannot detect the periodicity of $z(t)$ easily.

(d) **Periodic measure and Wasserstein distance from data**. In Figs. 3.3 and 3.4 we plot the periodic measures ρ_i, $i = 1, 2, \dots, 20$ for sub-dataset $\{y(i+20k)\}_{k=0}^{N-1}$, of which only approximations can be calculated using law of large numbers (Theorem 2.16). Here we take $N = 10{,}000$. The way to calculate the period of the periodic measure ρ_i, $i = 1, 2, \dots$, if its period is not known in advance, from the dataset which will be given later, concluded at the end of Example 3.2. We observe that the position of the bell curves moves from left-side to right-side and then back to left-side four times, and the bell shapes are nearly the same between ρ_i and ρ_{i+10} for $i = 1, 2, \dots, 10$, but different by a shift. These observations coincide with our analysis of the periodicity of the mean and that of the volatility from the probability distribution of $Y(t)$ being 4 and 10 respectively.

We calculate Wasserstein distance between ρ_1 and ρ_i which is denoted as $W(\rho_1, \rho_i)$ and plot them in Fig. 3.5 from the empirical approximation estimates of ρ_i for $i = 1, 2, \dots, 40$. Here ρ_i, $i = 21, 22, \dots, 40$ are computed from a dataset generated from the numerical scheme (3.3) with Brownian motion which is independent from the Brownian motion used to calculate $\rho_i, i = 1, 2, \dots, 20$. We can see that in one period the curve of $i \mapsto W(\rho_1, \rho_i)$ goes up and down roughly five times which corresponds to the mean trend but with some difference between each time which is due to the noise fluctuations. After one period the curve goes back to around zero as $W(\rho_1, \rho_{21}) \approx 7.0321 \times 10^{-4}$ for empirical approximations ρ_1 and ρ_{21}, from which we can conclude that $\rho_1 = \rho_{21}$. It is noted that our EPM scheme not only gives the periodicity of the correct period, but also provides desired distributions of the random periodic process at each time.

3.2. Detect an unknown period of a random periodic process

A challenging question here is that the true period q of the random periodic process or time series may be unknown to us. Thus we may not immediately be able to calculate the periodic measure ρ_i, $i = 1, 2, \dots$, as the law of large numbers to approximate ρ_i needs the true value of the period q. The law of large numbers along arithmetic progression of arbitrary common increment proved in the last section is significant in providing a statistical method using test periods to compute precisely the unknown period q for a dataset with random periodicity.

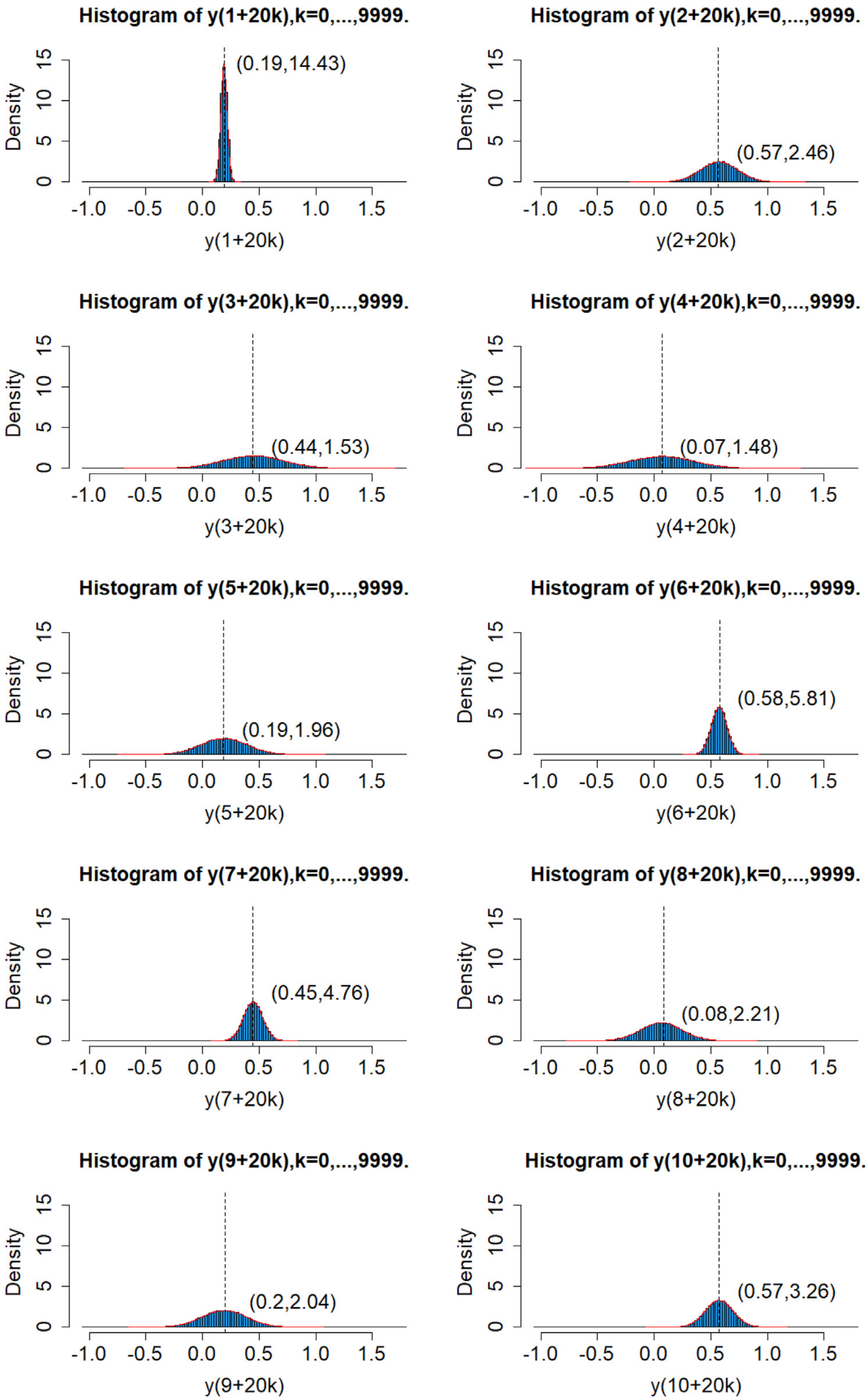


Fig. 3.3. Plots of histograms and density functions of periodic measures for $i = 1, \ldots, 10$.

The point is that the period, though may be unknown to us, exists in some sets of data. Our result gives a way to detect the period using test periods. Notice that $\mu^{i,p}$ depends on the test period p when the true period q is regarded as

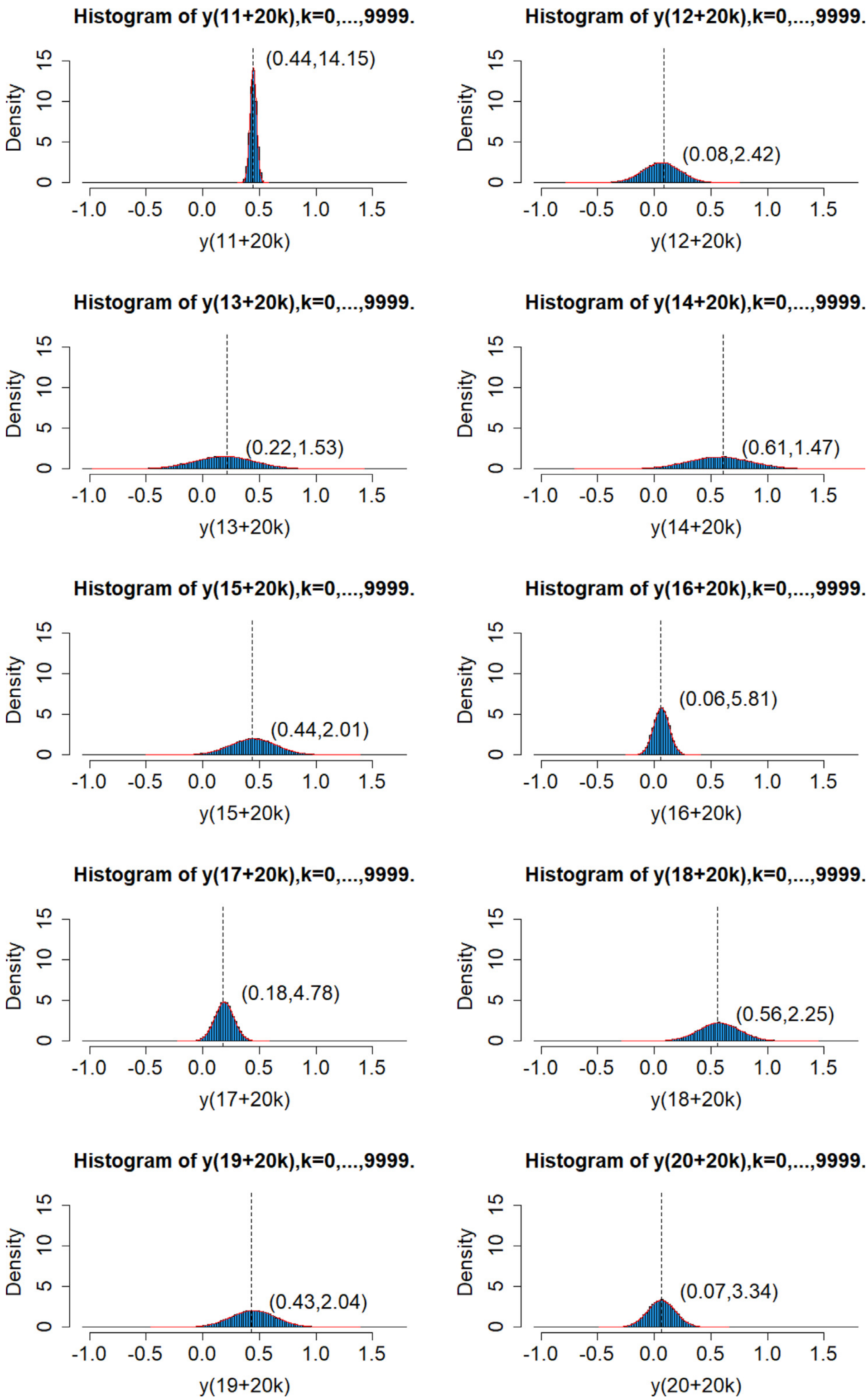


Fig. 3.4. Plots of histograms and density functions of periodic measures for $i = 11, \ldots, 20$.

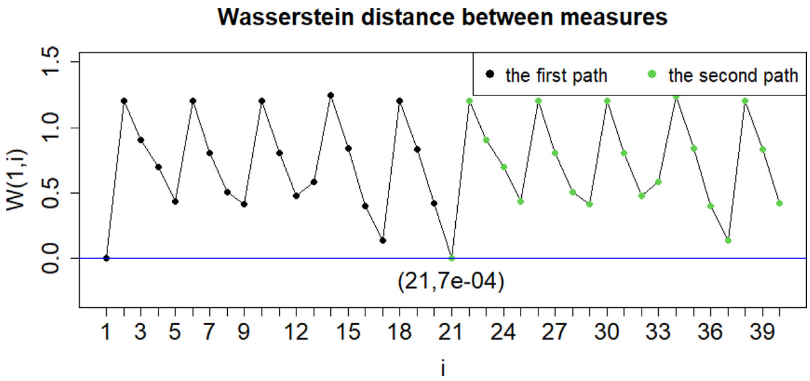


Fig. 3.5. Wasserstein distance $W(\rho_1, \rho_i)$, $i = 1, 2, \ldots, 40$.

fixed. We pick up p and find that $i \mapsto \mu^{i,p}$ has periodicity of period r, where r is the greatest common divisor of p and q. If it turns out that $i \mapsto \mu^{i,p}$ is aperiodic, then p, q are co-prime. If $i \mapsto \mu^{i,p}$ is periodic with period r, then r divides q. At $p = q$, $i \mapsto \mu^{i,p}$ has maximum period, in other words, if the period of $i \mapsto \mu^{i,p}$ is maximised at certain p, then $q = p$. In this context, the question remains to ask is: how do we know the period of $i \mapsto \mu^{i,p}$ is maximised at certain p? In the following, we provide a scheme to compute all the factors of such an integer.

We assume as a prior knowledge that $q \leq Q$ for certain integer Q. Note that any integer number can be decomposed as

$$q = r_1^{n_1} r_2^{n_2} \cdots r_m^{n_m}, \tag{3.4}$$

where $r_1 < r_2 \cdots < r_m$ are prime numbers and $n_1, n_2, \ldots, n_m$ are positive integers. If we can find all the prime number factors and their respective multipliers, we can find q. We start from the test period $p = 2$. For large N, consider the map of empirical measure approximation

$$i \mapsto \mu_K^{i,p} = \frac{1}{K} \sum_{k=0}^{K-1} I.(Y_{i+kp}(\cdot)). \tag{3.5}$$

If initially (3.5) is approximately an invariant measure, it means 2 or any power of 2 is not in the decomposition (3.4). If (3.5) appears to have period 2, then it means that 2 is a factor of q. Then we can continue to test $p = 2^2, 2^3, \ldots, 2^{r_1}$, for $r_1 \leq [\log_2 Q]$, then stop at one step $p = 2^{j_0+1}$ when 2^{j_0+1} is no longer the period of $i \mapsto \mu_K^{i,\cdot}$ (but 2^{j_0} is). In this case we know that 2^{j_0} is a factor of q and j_0 is the maximum power of factor $2^{\cdot}$ of the number q. We can decide any of the prime numbers and their powers appearing in the decomposition (3.4) by applying the above method for other possible p (noting the prior knowledge $q \leq Q$ here), thus eventually find the period q.

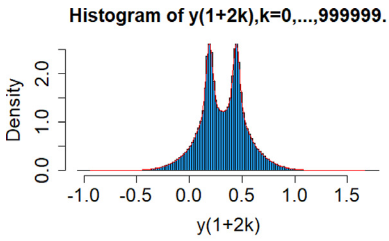
Example 3.2 (*Example 3.1 Continued*)**.** In this part we will study the empirical measures of sub-datasets $\{y(i+kp)\}_{k=0}^{K-1}$ for $i = 1, \ldots, p$ with different values of p.

Following the estimation scheme stated before, when $p = 2$, the dataset $\{y(t)\}_{t=1}^{n}$ are separated into two groups, one is that of the odd values of t and the other is that of the even values of t. We plot the histograms from the sub-datasets $\{y(i+kp)\}_{k=0}^{K-1}$ for $i = 1, 2$ in Fig. 3.6, $K = 100{,}000$ represents the length of data used to estimate the empirical measures. The figure suggests that the empirical measure of the odd valued group is different from that of the even valued group. Hence the map $i \mapsto \mu^{i,p}$ appears to have 2 different patterns.

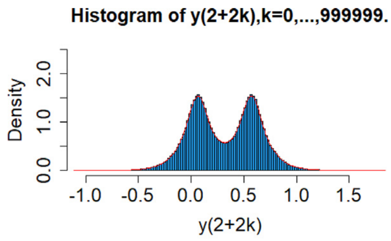
Set $\hat{W}_{(p)}(1, \cdot) : \mathbb{N} \to \mathbb{R}$ as the map $i \mapsto W(\mu_K^{1,p}, \mu_K^{i,p}) =: \hat{W}_{(p)}(1, i)$. To see the periodicity of Wasserstein distance $\hat{W}_{(2)}(1, \cdot)$, we simulate another path with the same initial value but under different noise and denote it as $\{\tilde{y}_t\}_{t=1}^{n}$. We denote the empirical measures of $\{\tilde{y}_{1+kp}\}_{k=0}^{K-1}$ and $\{\tilde{y}_{2+kp}\}_{k=0}^{K-1}$ as $\mu_K^{3,p}$ and $\mu_K^{4,p}$ respectively, and calculate $W(\mu_K^{1,p}, \mu_K^{i,p})$, $i = 3, 4$ (represented by green points) to compare with $W(\mu_K^{1,p}, \mu_K^{i,p})$, $i = 1, 2$ (represented in black points) in Fig. 3.6(c). This is to ensure the dataset used to compute $\mu_K^{i,p}$, $i = 1, 2$ is independent of the dataset to compute $\mu_K^{i,p}$, $i = 3, 4$. We see from the figure that these two series of values of Wasserstein distance are nearly the same, which suggests the period of $\hat{W}_{(2)}(1, \cdot)$ is 2.

We follow the same procedure as above to plot the limiting measures for $p = 2^2$ and $p = 2^3$ respectively. For $p = 2^2$, the limiting measures appear to have period 4 (see Fig. 3.7), while for $p = 2^3$, the limiting measures appear to have period 4 as well (see Figs. 3.8 and 3.9). From the discussion of this section, we know that 4 is a factor of the period q, but any number 2^i for $i \geq 3$ is not.

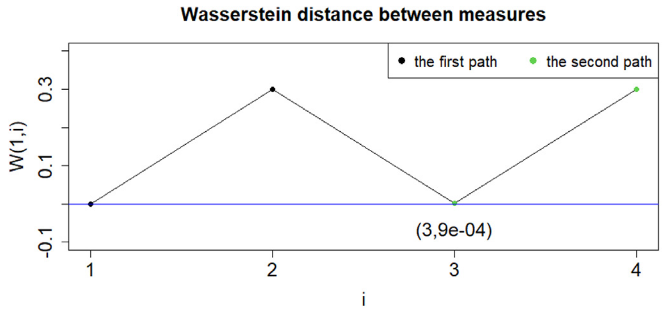
In the case that $p = 4$, there are 4 completely different sub-datasets and they generate 4 different histograms. In the case that $p = 8$, there are 8 completely different sub-datasets, but they still generate 4 different histograms.



(a) $\mu_K^{1,2}$



(b) $\mu_K^{2,2}$



(c) $i \mapsto W(\mu_K^{1,2}, \mu_K^{i,2}), i = 1, 2, 3, 4$

Fig. 3.6. Analysis of empirical measures of sub-datasets $\{y(i+kp)\}_{k=0}^{K-1}$ when $p = 2$.

The result in the case that $p = 8$ is remarkable if we notice the following observation. Take $i = 1$, then the sub-dataset $\{y_1, y_9, y_{17}, \ldots\}$ is used to generate $\mu^{1,8}(\cdot)$ and the sub-dataset $\{y_5, y_{13}, y_{21}, \ldots\}$ is used to generate $\mu^{5,8}(\cdot)$. Note the datasets $\{y_1, y_9, y_{17}, \ldots\}$ and $\{y_5, y_{13}, y_{21}, \ldots\}$ have no intersections at all, but they produce the same empirical measure $\mu^{1,8}(\cdot) = \mu^{5,8}(\cdot)$. This agrees entirely with Theorem 2.16 that we proved in the last section. Same remark applies for $\mu^{2,8} = \mu^{6,8}$, $\mu^{3,8} = \mu^{7,8}$ and $\mu^{4,8} = \mu^{8,8}$. With this remark, it is not hard to understand why it is reliable that we use two different datasets generated by two independent noises to compute $\mu_K^{1,2}$ and $\mu_K^{3,2}$. Same reason applies to $\mu_K^{1,4}$ and $\mu_K^{5,4}$.

Next to test whether 3 is a prime factor of q, we plot histograms from the datasets $\{y(i+3k)\}_{k=0}^{K-1}$ for $i = 1, 2, 3$ in Fig. 3.10(a) to (c). Here $K = [2000000/3] = 666,666$. All the histograms are nearly the same. Besides, there is no periodicity in Fig. 3.10(d) and the Wasserstein distances $W(\mu_K^{1,3}, \mu_K^{i,3})$ are very closed to 0 for all $i = 1, 2, \ldots, 6$. It is noted here that the empirical measures appear to have period 1, which shows that $p = 3$ and q are co-prime and thus 3 is not a factor of q.

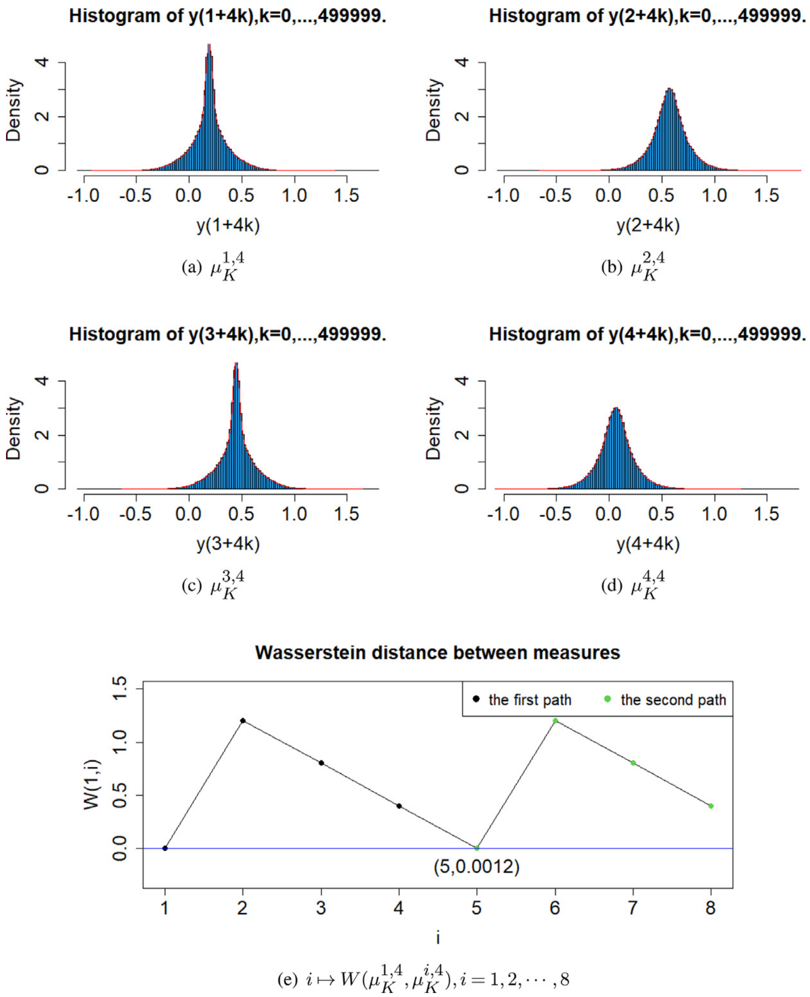
Again it is interesting to note when $i = 1, 2, 3$, the three sub-datasets $\{y_1, y_4, y_7, \ldots\}$, $\{y_2, y_5, y_8, \ldots\}$ and $\{y_3, y_6, y_9, \ldots\}$ are completely different, but they generate the identical empirical probability measures. The histograms generated by three completely different datasets are nearly the same due to that the length of the dataset is large enough. To make it more clear we also show the figure of Wasserstein distance $i \mapsto W(\mu_K^{1,9}, \mu_K^{i,9})$ when $p = 9$ in Fig. 3.11(a).

Then we test the prime number 5 and show the figures of the Wasserstein distances from the sub-datasets $\{y(i+kp)\}_{k=0}^{K-1}$ for $i = 1, 2, \ldots, p$ with $p = 5$ and $p = 25$ respectively in Fig. 3.11(b) and (c). Here we take $K = 40{,}000$ when $p = 5$ and $K = 8000$ when $p = 25$. The curves of $i \mapsto W(\mu_K^{1,p}, \mu_K^{i,p})$ show clear periodicity with period 5 for both $p = 5$ and $p = 25$. Similar analysis as above shows that $r_2 = 5$ is another factor of q and its power is $n_2 = 1$.

Now we have found both 4 and 5 are factors of the true period q, so 20 must divide q. Thus if we take $p = 20$, then $i \mapsto \mu^{i,20}$ should be a periodic function with period 20. Indeed, following our numerical scheme and by similar analysis, we find that when $p = 20$ the empirical measures $\mu_K^{i,20}$ from $\{y(i+kp)\}_{k=0}^{K-1}$ for $i = 1, 2, \ldots, 20$ appear to have period 20. See Figs. 3.3–3.5.

Let us now suppose the actual value of q is not known, but we already concluded that 20 can divide q. Assume we also have other knowledge about q, e.g. if we can ensure that $q \le 1000$. In order to know the exact value of q, we can test prime numbers $p \le 50$: 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47 respectively. In fact, we find from our dataset that none of them is a factor of q. The details are omitted.

Having known 20 is a factor of q, 2^3 and 5^2 are not, so any multiplication of 20 and 2^i, $i \ge 1$, or 5^i, $i \ge 1$, is not a factor of q, and the prime numbers 3, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47 are not factors of q, then we can conclude the biggest possible factor of q, which is less than or equal to 1000, is 20. For this we used the fact that 50 cannot be a factor of q as we know 5^2 is not a factor of q. Thus we conclude that $q = 20$.



(a) $\mu_K^{1,4}$ (b) $\mu_K^{2,4}$

(c) $\mu_K^{3,4}$ (d) $\mu_K^{4,4}$

(e) $i \mapsto W(\mu_K^{1,4}, \mu_K^{i,4}), i = 1, 2, \cdots, 8$

Fig. 3.7. Analysis of empirical measures of sub-datasets $\{y(i + kp)\}_{k=0}^{K-1}$ when $p = 4$.

3.3. An example with strong noise

The following example has strong noise for which we will see that DFT does not work robustly, but the EPM scheme still works well.

Example 3.3. Consider the following SDE,

$$dX(t) = \left(-\pi X(t) + 0.1 \sin\left(\frac{\pi t}{2}\right) + 1\right) dt + \left(0.1 + 10 \sin\left(\frac{\pi t}{5}\right)\right) dW_t, \tag{3.6}$$

with $X(0) = x$. In this example, the periodicity of mean is weak and the noise perturbation is strong.

By a similar calculation we know that the periodic measure ρ_t of SDE (3.6) is a Gaussian distribution with mean $\frac{1}{\pi} + \frac{1}{80}\left(\frac{\sin(\frac{\pi t}{2})}{\pi} - \frac{\cos(\frac{\pi t}{2})}{2\pi}\right)$ and variance $\frac{0.1^2}{2\pi} + (2 \times 0.1 \times 10) \times \frac{100}{101}\left(\frac{\sin(\frac{\pi t}{5})}{2\pi} - \frac{\cos(\frac{\pi t}{5})}{20\pi}\right) + \frac{10^2}{2}\left(\frac{1}{2\pi} - \frac{25}{26}\left(\frac{\cos(\frac{2\pi t}{5})}{2\pi} + \frac{\sin(\frac{2\pi t}{5})}{10\pi}\right)\right)$. It has period 20 as well. The variance is comparatively larger than the variation of the mean with respect to time t.

We find from this example that the larger the noise is, the more data DFT needs in order to obtain stable results. We compare the periodogram for different datasizes: $n_1 = 10^4$, $n_2 = 10^5$, $n_3 = 2 \times 10^5$, $n_4 = 4 \times 10^5$, $n_5 = 4.4 \times 10^5$ and $n_6 = 5 \times 10^5$. We observe that if the length of the dataset is less than around 4.5×10^5, the dominated period obtained from DFT varies with different n, but none of them is even 4. See Fig. 3.12. It seems that the DFT method cannot work robustly in this situation. For any datasize, there is no way for DFT to detect the true period $q = 20$ in this example.

To see the efficiency of the EPM scheme, we compute the periodic measures ρ_i and the Wasserstein distance $W(\rho_1, \rho_i)$ for $i = 1, \ldots, 40$ with datasize $n = 2 \times 10^6$. The later is plotted in Fig. 3.13(a). It looks as if that the mapping $i \mapsto \rho_i$ and

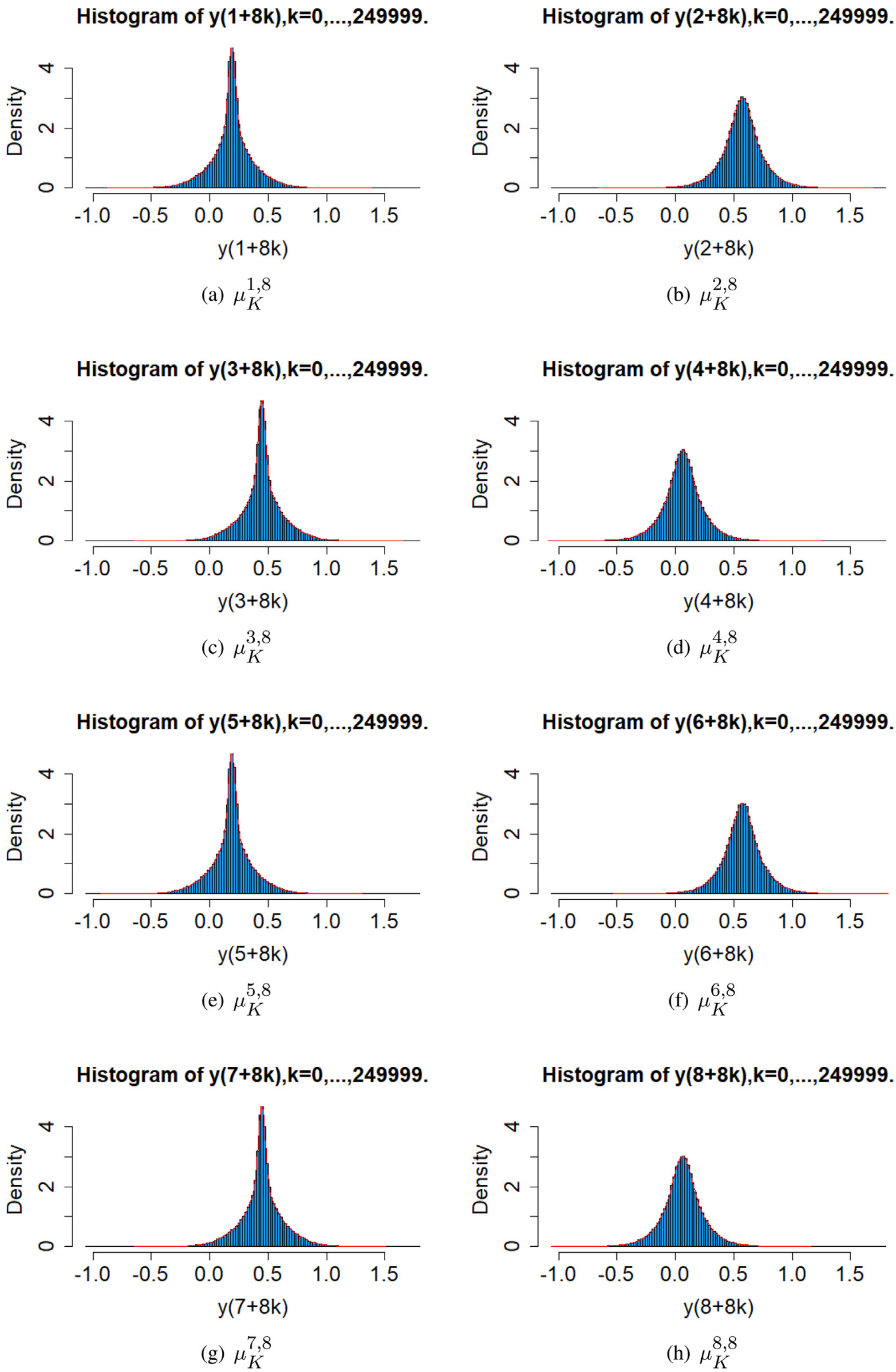


Fig. 3.8. Analysis of measures of sub-datasets $\{y(i+kp)\}_{k=0}^{K-1}$ when $p=8$.

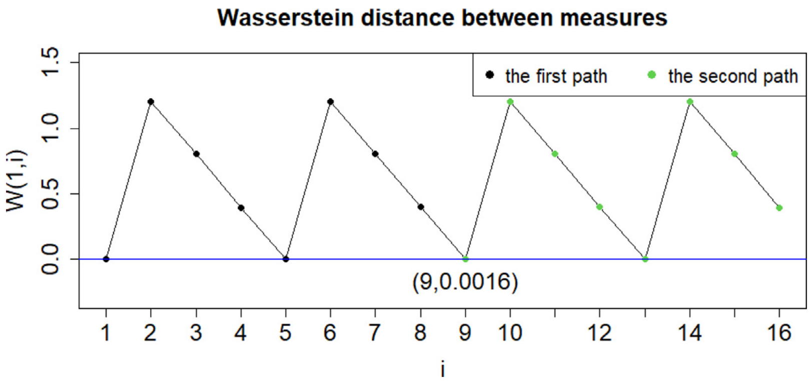
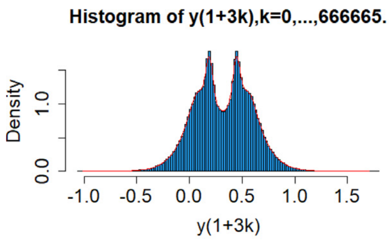
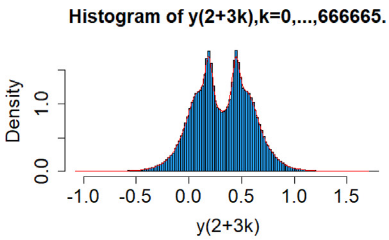


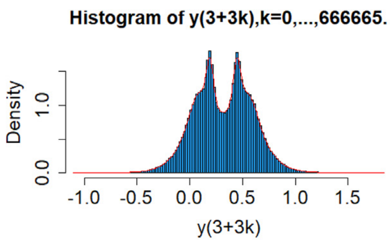
Fig. 3.9. $i \mapsto W(\mu_K^{1,8}, \mu_K^{i,8})$, $i = 1, 2, \ldots, 16$.



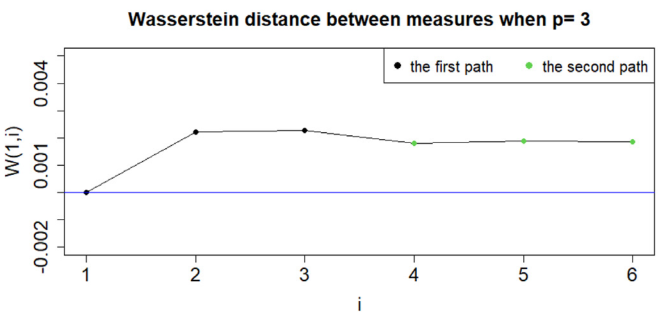
(a) $\mu_K^{1,3}$



(b) $\mu_K^{2,3}$

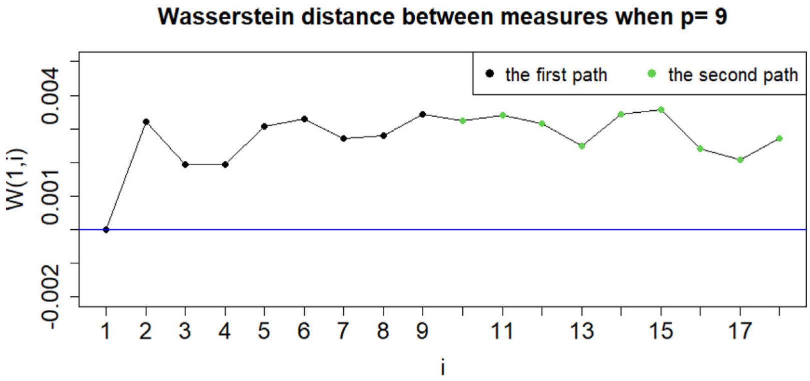


(c) $\mu_K^{3,3}$

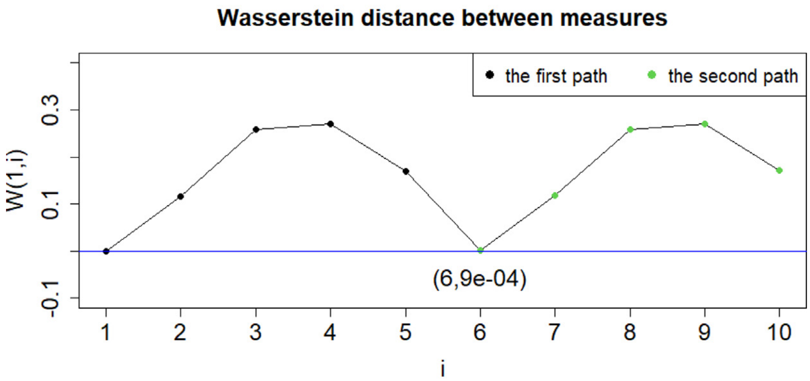


(d) $i \mapsto W(\mu_K^{1,3}, \mu_K^{i,3}), i = 1, 2, \cdots, 6$

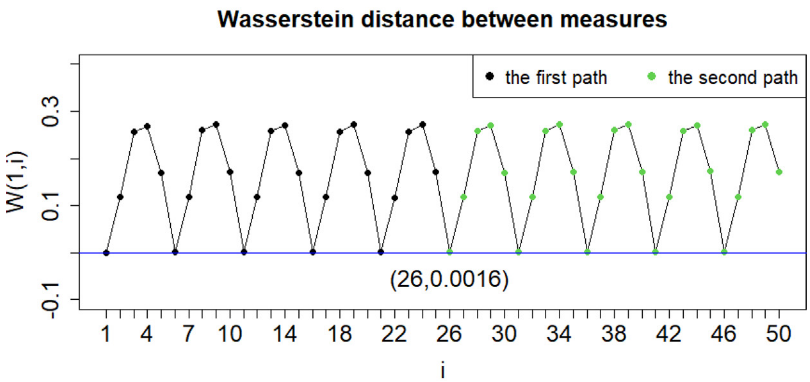
Fig. 3.10. Analysis of measures of sub-datasets $\{y(i + kp)\}_{k=0}^{K-1}$ when $p = 3$.



(a) $i \mapsto W(\mu_K^{1,9}, \mu_K^{i,9}), i = 1, 2, \cdots, 18$



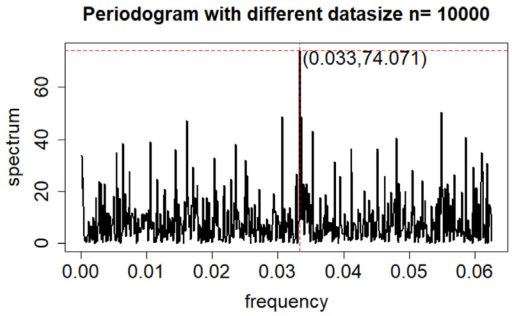
(b) $i \mapsto W(\mu_K^{1,5}, \mu_K^{i,5}), i = 1, 2, \cdots, 10$



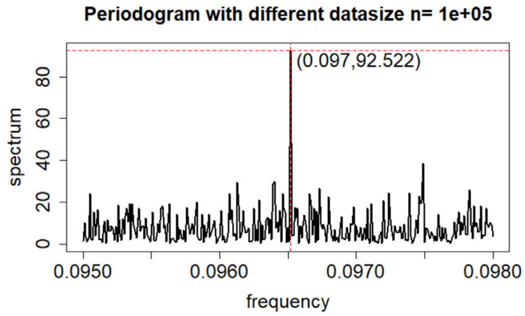
(c) $i \mapsto W(\mu_K^{1,25}, \mu_K^{i,25}), i = 1, 2, \cdots, 50$

Fig. 3.11. The Wasserstein distances $i \mapsto W(\mu_K^{1,p}, \mu_K^{i,p})$ for different $p = 5, 9, 25$.

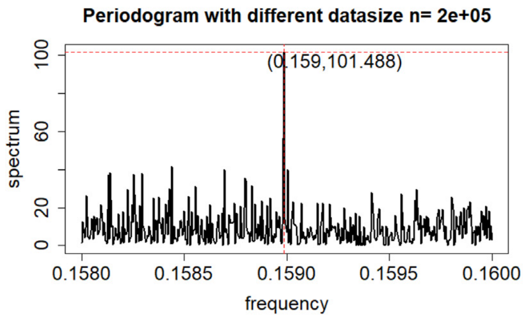
$i \mapsto W(\rho_1, \rho_i)$ had period 5. In fact, we can see in Fig. 3.13(b) the values of $W(\rho_1, \rho_6) \approx 0.0555$, $W(\rho_1, \rho_{11}) \approx 0.02538$, $W(\rho_1, \rho_{16}) \approx 0.0493$, $W(\rho_1, \rho_{26}) \approx 0.0567$, $W(\rho_1, \rho_{31}) \approx 0.0245$, $W(\rho_1, \rho_{36}) \approx 0.0481$. Though they are quite small comparing with Wasserstein distance of other i (except for $i = 21$), $W(\rho_1, \rho_{21}) \approx 0.00202$ is 10 times smaller than them. Thus it is reasonable to say that the period is 20. Note here $\rho_i, i = 21, 22, \ldots, 40$ are computed from a dataset generated from a Brownian motion which is independent from the Brownian motion used to generate $\rho_i, i = 1, 2, \ldots, 20$.



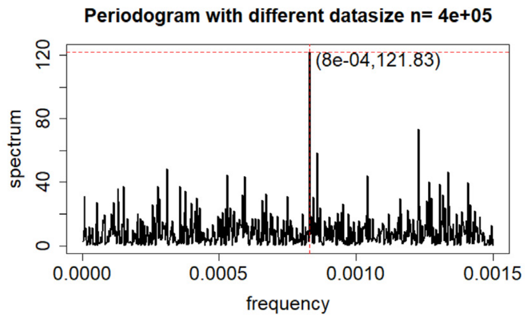
(a) Periodogram of datasize $n_1 = 10^4$: $f_{\max} = 0.0\dot{3}$ and the corresponding dominated period is 30.



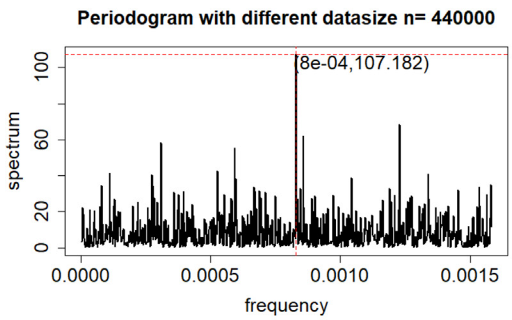
(b) Periodogram of datasize $n_2 = 10^5$: $f_{\max} = 0.09652$ and the corresponding dominated period is 10.36.



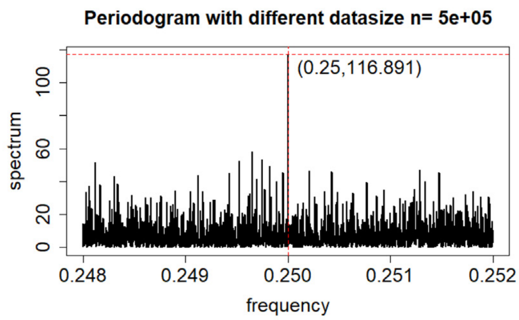
(c) Periodogram of datasize $n_3 = 2 \times 10^5$: $f_{\max} = 0.158985$ and the corresponding dominated period is 6.29.



(d) Periodogram of datasize $n_4 = 4 \times 10^5$: $f_{\max} = 0.00083$ and the corresponding dominated period is 1204.82.



(e) Periodogram of datasize $n_5 = 4.4 \times 10^5$: $f_{\max} \approx 0.00083$ and the corresponding dominated period is 1205.36.

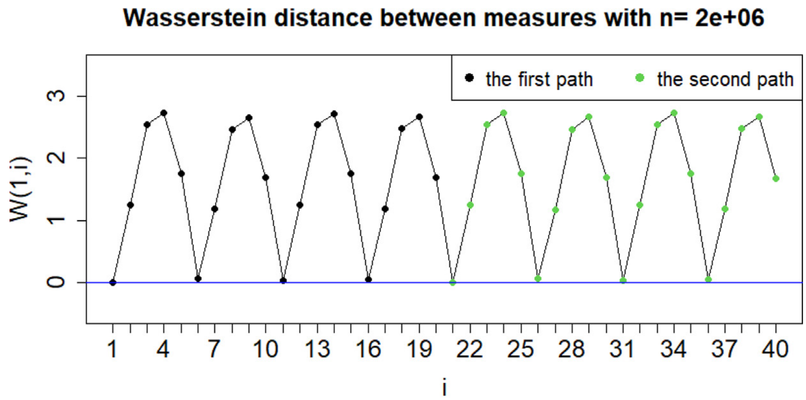


(f) Periodogram of datasize $n_6 = 5 \times 10^5$: $f_{\max} = 0.025$ and the corresponding dominated period is 4.

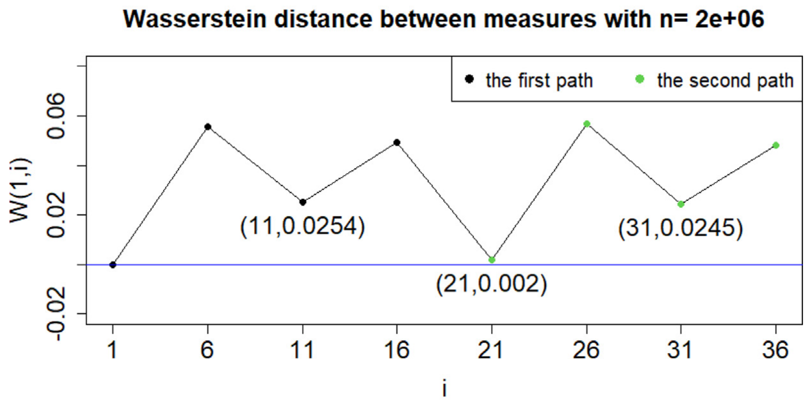
Fig. 3.12. Periodograms of datasets $\{y(t)\}_{t=1}^{n}$ for different datasize n.

To verify the robustness of the EPM scheme for different amount of data, we calculate the Wasserstein distance $i \mapsto W(\mu_K^{1,20}, \mu_K^{21,20})$ for different datasizes, $n_1 = 10^4$, $n_2 = 10^5$, $n_3 = 2 \times 10^5$, $n_5 = 4.4 \times 10^5$, $n_6 = 5 \times 10^5$ and $n_7 = 2 \times 10^6$, and plot the curve of $n \mapsto W(\mu_K^{1,20}, \mu_K^{21,20})$ $(K = [n/p])$ in Fig. 3.14. We can see that the Wasserstein distance $W(\mu_K^{1,20}, \mu_K^{21,20})$, except for the case that $n_1 = 10^4$, is closed to zero for all other bigger datasizes $n_2, \ldots, n_7$, and the general pattern of $i \mapsto W(\rho_1, \rho_i)$ keeps consistent for different datasizes. This shows that the EPM scheme is robust.

From Fig. 3.14 we can see that the error of $W(\mu_K^{1,20}, \mu_K^{21,20}) < 0.01$ when $n > 10^5$. So we can obtain relatively accurate result about the real period q with less data by using the EPM scheme comparing with the DFT method.



(a) $W(\rho_1, \rho_i), i = 1, \cdots, 40$



(b) $W(\rho_1, \rho_i), i = 1, 6, 11, 16, 21, 26, 31$

Fig. 3.13. Figures of the Wasserstein distance $W(\rho_1, \rho_i)$ for $i = 1, \ldots, 40$.

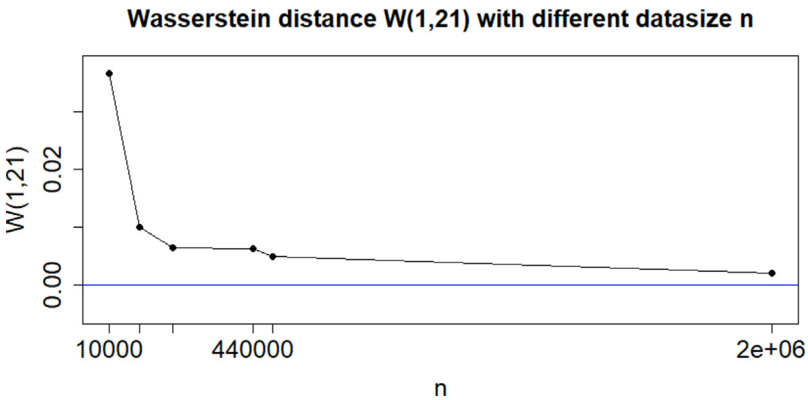


Fig. 3.14. Wasserstein distance $W(\rho_1, \rho_{21})$ with different datasize n.

3.4. A real world example

Example 3.4. Recent work about modelling real-world temperature and the weather derivative market uses periodic function to estimate the volatility of temperature dataset. This gives rise to our interest to apply our EPM scheme of finding its period and distributions. Relevant references include [33–37].

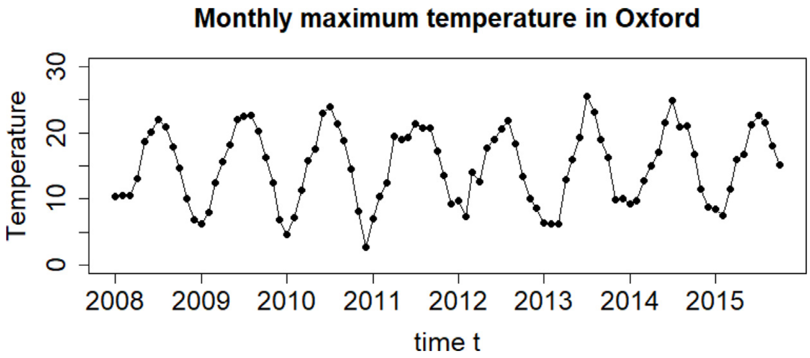


Fig. 3.15. Daily maximum temperature in Central England.

The dataset $\{y(t)\}$ we used in this example is the monthly average maximal temperature in Oxford. It contains 1954 monthly data for almost 163 years from Jan. 1853 to Oct. 2015, which are computed from the daily maximal temperature records. Part of the data is plotted in Fig. 3.15.

Suppose we do not know the exact value of q, following the estimation scheme, we can find $q = r_1^{n_1} r_2^{n_2}$ with $r_1 = 2, r_2 = 3$ and $n_1 = 2, n_2 = 1$. We refer to Figs. 3.16 and 3.17 from the periodic measure ρ_i with period 12. We can obtain information about the mean and volatility for each month from the empirical periodic measures. For example, we can see that the volatility of daily maximal temperature in the winter is relatively larger than that in the summer in Oxford.

To plot the Wasserstein distance between ρ_1 and ρ_i, we separate the dataset into two groups, one is of Year 1878–1958 and the other is of Year 1959–2015. We use the first group to estimate $\rho_1, \ldots, \rho_{12}$ (represented by black spots), and use the second one to estimate $\rho_{13}, \ldots, \rho_{24}$ (represented by green spots), and plot the Wasserstein distance $W(\rho_1, \rho_i)$ in Fig. 3.18. We can see that the curve $i \mapsto W(\rho_1, \rho_i)$ appears to have periodic pattern with period 12. But we should point out that $W(\rho_1, \rho_{13}) \approx 0.016637$ is not very closed to zero as we may have hoped. This may be because the size of data is not large enough.

Remark 3.5. We have seen that it is very efficient to apply Theorems 2.16 and 2.20 to datasets of time series. Following Remark 2.17 and Theorem 2.18, the assumption in Theorem 2.16 that the discrete metric dynamical system is ergodic holds naturally if the random dynamical system given by the time series is driven by a Brownian motion. Therefore in order to apply Theorems 2.16 and 2.20, we only need to use some standard statistical methods to test the normality of noise in the dataset to guarantee our theory is valid. For references about the normality test see [38–40].

For instance, let us consider Example 3.4. We first use $s(t) = Bt + Ccos(\frac{2\pi}{q}t + \phi) + D$ to estimate both the non-periodic trend and the seasonal trend of dataset and remove it. By a standard least-square method we can find the parameters as

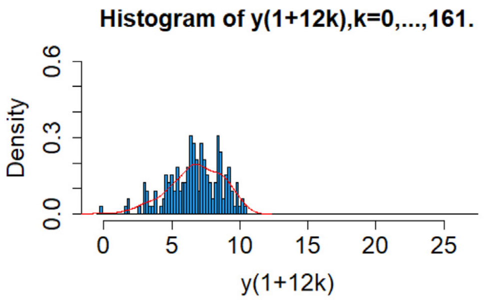
$$B = 6.282 \times 10^{-4}, \quad C = 7.6359, \quad D = 13.2968, \quad \phi = 5.6971.$$

De-trended dataset can then be obtained by deducing the deterministic function from the data and its normality can be checked by a variety of different methods. Common methods of normality test in statistics are QQ-plot, histogram and Kolmogorov–Smirnov test. The QQ-plot and histogram give us a rough and visualised impression about the normality of dataset. If the tested dataset is normal, then the QQ-plot will coincide with the line and the histogram and estimated density function will coincide with normal density function (see Fig. 3.19).

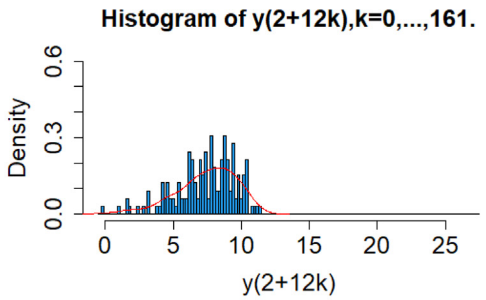
We can do a Kolmogorov–Smirnov Goodness-of-Fit test to test the null hypothesis H_0 that the tested sample comes from the distribution F_0 (here we take F_0 as norm distribution). The test statistic is $D_n = \sup_x |F_n(x) - F_0(x)|$, where n is the size of dataset and $F_n(x)$ is the empirical cumulative distribution function at x computed from dataset (see [39]). The critical value of dataset with size $n = 1954$ is $D_c \approx 0.0307$. The p-value of a hypothesis test tells how likely it is that the dataset could have occurred under the null hypothesis by calculating the likelihood of the test statistic. If D_n is smaller than the critical value D_c and the p-value is greater than the given alpha level, then we accept the null hypothesis. The result of Kolmogorov–Smirnov test of this example is: $D_n = 0.017346$, the corresponding p-value = 0.599. Since the p-value is greater than 0.05, we do not reject the null hypothesis, we can claim that the noise is Gaussian white noise in time.

Conclusion

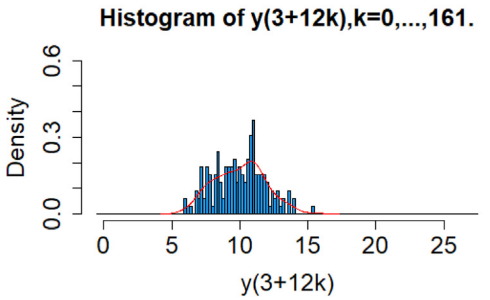
We have established an innovative ergodic periodic measure (EPM) approach to analyse the periodicity of a time series dataset using the recently developed ergodic theory of random periodic processes as an alternative to the Discrete Transformation method. The EPM scheme works efficiently and has clear advantages to be able to detect periodicity of



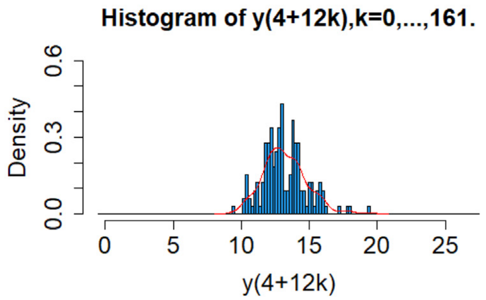
(a) ρ_1



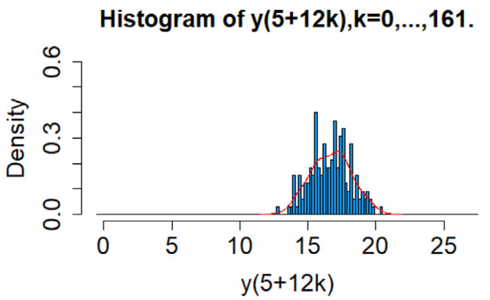
(b) ρ_2



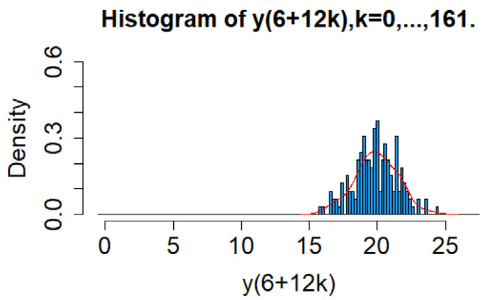
(c) ρ_3



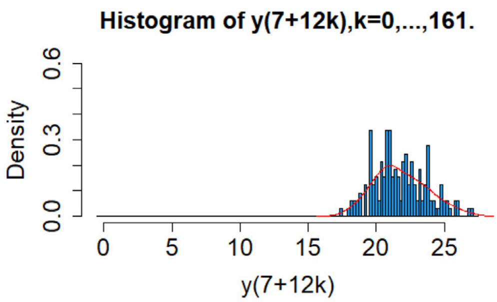
(d) ρ_4



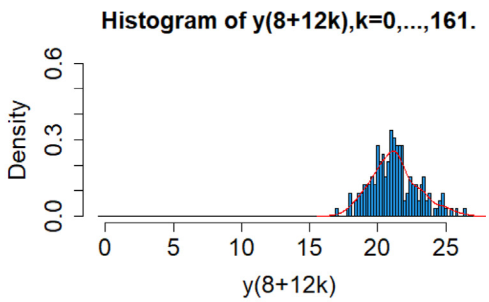
(e) ρ_5



(f) ρ_6

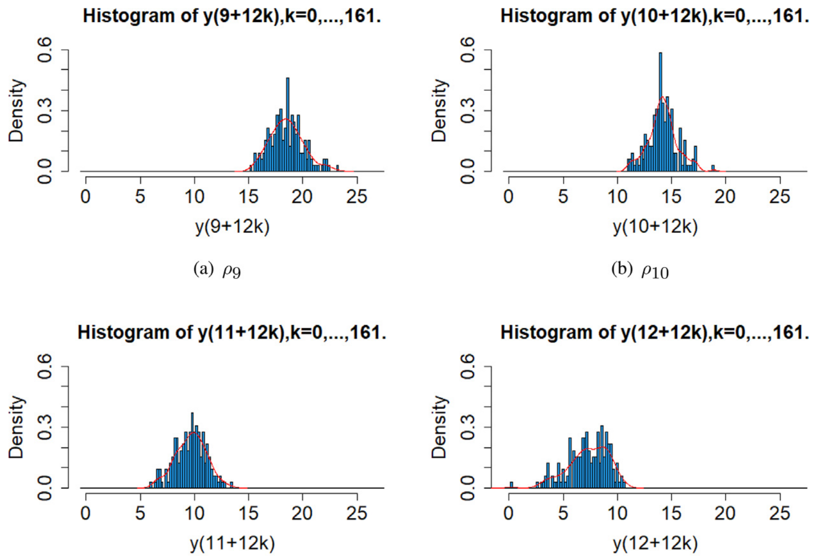


(g) ρ_7



(h) ρ_8

Fig. 3.16. Histograms of empirical periodic measures of monthly maximum temperature in Central England for months from January to October.



(a) ρ_9 (b) ρ_{10}

(c) ρ_{11} (d) ρ_{12}

Fig. 3.17. Histograms of empirical periodic measures of monthly maximum temperature for months from September to December.

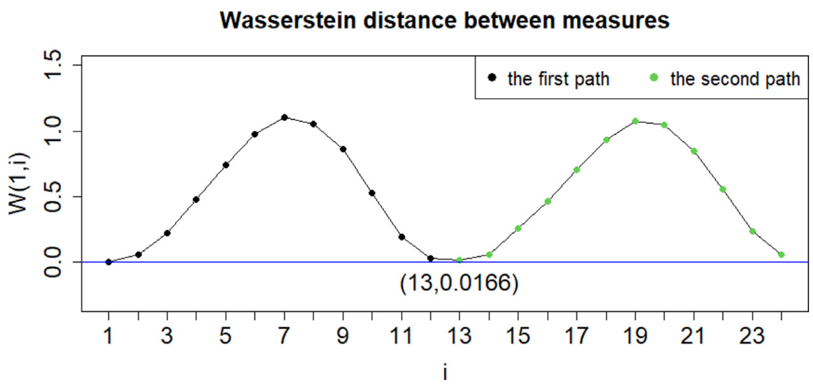


Fig. 3.18. Wasserstein distance $i \mapsto W(\rho_1, \rho_i)$ for $i = 1, \ldots, 24$.

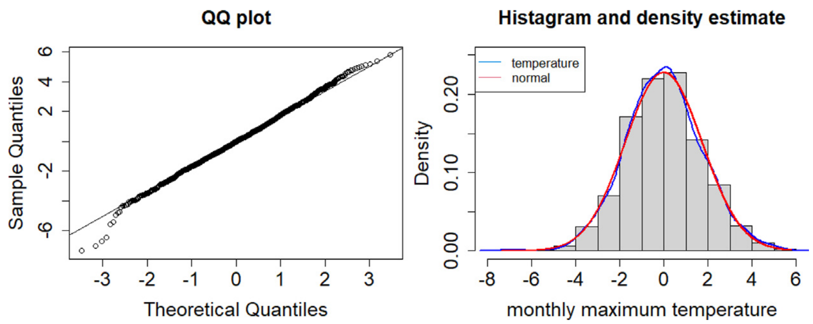


Fig. 3.19. QQ-plot and histogram for de-trended path of monthly maximum temperature dataset.

distributions and to compute the period over the classical Fourier series approach. In some situations, while the latter method cannot work robustly, our method can still work well. In this paper, we have built the probability foundation for this approach. The key idea is the law of large numbers for a test period. This makes it possible to use test periods to analyse the periodicity of a time series and detect its period by analysing the periodicity of empirical distributions out of a subsequence of a time series for the test periods. It was also observed that the periodicity of empirical distributions can be quantified by the Wasserstein distance. The latter can be computed from datasets. However, it looks that to achieve great accuracy, the amount of data still needs to be large in the case especially when noise is strong, so some ideas to reduce the amount of data needed are very much worth exploring further in the future, especially for applications.

CRediT authorship contribution statement

Chunrong Feng: Conceptualization, Methodology, Writing, Review & editing. **Yujia Liu:** Conceptualization, Methodology, Writing, Numerical simulations. **Huaizhong Zhao:** Conceptualization, Methodology, Writing, Review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

We are grateful to David Elworthy at the University of Warwick and Yu Liu at Durham University for useful conversations and comments. We would like to thank the anonymous referee for their constructive comments which have resulted in significant improvements of the paper. We acknowledge the financial supports of an EPSRC Established Career Fellowship to HZ (ref. EP/S005293/2) and a Royal Society grant (ref. NIF\R1\221003).

References

[1] Schuster A. On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena. Terr. Magn. 1898;3(1):13–41.
[2] Zhao H, Zheng Z-H. Random periodic solutions of random dynamical systems. J Differential Equations 2009;246(5):2020–38.
[3] Feng C, Zhao H. Random periodic processes, periodic measures and ergodicity. J Differential Equations 2020;269(9):7382–428.
[4] Feng C, Zhao H, Zhou B. Pathwise random periodic solutions of stochastic differential equations. J Differential Equations 2011;251(1):119–49.
[5] Feng C, Qu B, Zhao H. Random quasi-periodic paths and quasi-periodic measures of stochastic differential equations. J Differential Equations 2021;286:119–63.
[6] Wang B. Existence, stability and bifurcation of random complete and periodic solutions of stochastic parabolic equations. Nonlinear Anal 2014;103:9–25.
[7] Bates PW, Lu K, Wang B. Attractors of non-autonomous stochastic lattice systems in weighted spaces. Physica D 2014;289:32–50.
[8] Cherubini AM, Lamb JSW, Rasmussen M, Sato Y. A random dynamical systems perspective on stochastic resonance. Nonlinearity 2017;30(7):2835–53.
[9] Feng C, Zhao H, Zhong J. Existence of geometric ergodic periodic measures of stochastic differential equations. J Differential Equations 2023;(in press).
[10] Feng C, Zhao H, Zhong J. Expected exit time for time-periodic stochastic differential equations and applications to stochastic resonance. Physica D 2021;417:18, Paper No. 132815.
[11] Feng C, Liu Y, Zhao H. Ergodic numerical approximation to periodic measures of stochastic differential equations. J Comput Appl Math 2021;398:23, Paper No. 113701.
[12] Huang W, Lian Z, Lu K. Ergodic theory of random Anosov systems mixing on fibers. 2016, arXiv preprint arXiv:1612.08394.
[13] Chekroun MD, Simonnet E, Ghil M. Stochastic climate dynamics: random attractors and time-dependent invariant measures. Physica D 2011;240(21):1685–700.
[14] Engel M, Kuehn C. A random dynamical systems perspective on isochronicity for stochastic oscillations. Comm Math Phys 2021;386(3):1603–41.
[15] Gao P, Liu Y, Sun Y, Zheng Z. Large deviations principle for stationary solutions of stochastic differential equations with multiplicative noise. 2022, arXiv preprint arXiv:2206.02356v1.
[16] Branicki M, Uda K. Time-periodic measures, random periodic orbits, and the linear response for dissipative non-autonomous stochastic differential equations. Res Math Sci 2021;8:42–62.
[17] Uda K. Averaging principle for stochastic differential equations in the random periodic regime. Stochastic Process Appl 2021;139:1–36.
[18] Cheban D, Liu Z. Periodic, quasi-periodic, almost periodic, almost automorphic, Birkhoff recurrent and Poisson stable solutions for stochastic differential equations. J Differential Equations 2020;269(4):3652–85.
[19] Raynaud de Fitte P. Almost periodicity and periodicity for nonautonomous random dynamical systems. Stoch Dyn 2021;21(06):2150034.
[20] Gao L, Yan L. On random periodic solution to a neutral stochastic functional differential equation. Math Probl Eng 2018;8353065, 9.
[21] Dong Z, Zheng Z, Zhang W. Random periodic solutions of non-autonomous stochastic differential equations. 2021, arXiv preprint arXiv:2104.01423.
[22] Liu R, Lu K. Statistical properties of 2D stochastic Navier-Stokes equations with time-periodic forcing and degenerate stochastic forcing. 2021, arXiv preprint arXiv:2105.00598.
[23] Song J, Song X, Zhang Q. Nonlinear feynman-kac formulas for stochastic partial differential equations with space-time noise. SIAM J Math Anal 2019;51(2):955–90.
[24] Birkhoff G. Proof of the ergodic theorem. Proc Natl Acad Sci USA 1932;17:656–60.
[25] Elworthy KD. Stochastic differential equations on manifolds. London mathematical society lecture note series, vol. 70, Cambridge-New York: Cambridge University Press; 1982, p. xiii+326.
[26] Silverman JH. A friendly introduction to number theory. Brooklyn, New York: Pearson Education Inc.; 2014.

[27] Feng C, Qu B, Zhao H. A sufficient and necessary condition of PS-ergodicity of periodic measures and generated ergodic upper expectations. Nonlinearity 2020;33(10):5324–54.
[28] Fournier N, Guillin A. On the rate of convergence in Wasserstein distance of the empirical measure. Probab Theory Related Fields 2015;162(3–4):707–38.
[29] Bloomfield P. Fourier analysis of time series: an introduction. New York: Wiley; 1976.
[30] Panaretos VM, Zemel Y. Statistical aspects of Wasserstein distances. Annu Rev Stat Appl 2019;6:405–31.
[31] Arnold L. Random dynamical systems. Springer monographs in mathematics, Berlin: Springer-Verlag; 1998, p. xvi+586.
[32] Feng C, Liu Y, Zhao H. Numerical approximation of random periodic solutions of stochastic differential equations. Z Angew Math Phys 2017;68(5).
[33] Alaton P, Djehiche B, Stillberger D. On modelling and pricing weather derivatives. Appl Math Finance 2002;9(1):1–20.
[34] Benth F, Saltyte Benth J. Stochastic modelling of temperature variations with a view towards weather derivatives. Appl Math Finance 2005;12:53–85.
[35] Benth F, Saltyte Benth J. The volatility of temperature and pricing of weather derivatives. Quant Finance 2007;7:553–61.
[36] Brody DC, Syroka J, Zervos M. Dynamical pricing of weather derivatives. Quant Finance 2002;2:189–98.
[37] Campbell S, Diebold F. Weather forecasting for weather derivatives. J Amer Statist Assoc 2005;100:6–16.
[38] Chambers J, Cleveland W, Kleiner B, Tukey P. Graphical methods for data analysis, Vol. 40. Belmont, Calif: Wadsworth International Group; 1983.
[39] Kolmogorov AN. Sulla determinazione empirica di una legge di distribuzione. Giorn. Ist. Ital. Attuari 1933;4:83–91.
[40] Wilk MB, Gnanadesikan R. Probability plotting methods for the analysis for the analysis of data. Biometrika 1968;55(1):1–17.