Can Arrival Rates be Modelled by Sine Waves?

Ningyuan Chen[1], joint work with Donald Lee[2] and Sahand Negaban[3]

Rotman Healthcare Roundtable 2020

[1]UToronto
[2]Emory
[3]Yale

INTRODUCTION

Background

- Arrival processes describe temporal demand for service in queuing systems. It is the starting point of all subsequent operations.

Background

- Arrival processes describe temporal demand for service in queuing systems. It is the starting point of all subsequent operations.
- Emergency departments:
    - Capacity and staffing plans require a good understanding of patient arrival patterns.
    - Poor forecasting of demand can rob patients of timely critical care.

Background

- Arrival processes describe temporal demand for service in queuing systems. It is the starting point of all subsequent operations.
- Emergency departments:
  - Capacity and staffing plans require a good understanding of patient arrival patterns.
  - Poor forecasting of demand can rob patients of timely critical care.

- Many other examples where accurate models for arrivals are critical to managers

Common Practice

- How are arrival rates specified or estimated in practice from time stamps of past arrivals?
  - Specify a period (say, a week) such that the arrival pattern repeats itself judged from experience

Common Practice

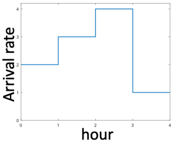
- How are arrival rates specified or estimated in practice from time stamps of past arrivals?
  - Specify a period (say, a week) such that the arrival pattern repeats itself judged from experience
  - Specify a bucket size (say, an hour) and count the arrivals in each bucket

COMMON PRACTICE

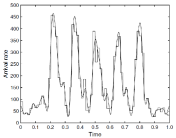
- How are arrival rates specified or estimated in practice from time stamps of past arrivals?
  - Specify a period (say, a week) such that the arrival pattern repeats itself judged from experience
  - Specify a bucket size (say, an hour) and count the arrivals in each bucket
  - Average the bucket counts across periods
  - (Optional) fit the piecewise constant curve by a function

Time of arrival (hr)
0.34
0.45
1.2
1.8
1.88
2.04
2.05
2.08
2.5
3.6

Hr	Arr. per hr
1st	2
2nd	3
3rd	4
4th	1



Example Simple piecewise constant fit



Example Email arrivals over a wk – smooth fit using 50 functions

Strengths and Weaknesses

- Strengths
    - Robust
    - No need to specify a model (nonparametric)
    - Efficient to compute

Strengths and Weaknesses

- Strengths
  - Robust
  - No need to specify a model (nonparametric)
  - Efficient to compute
- Weaknesses
  - Prior knowledge of the frequency
  - Cannot deal with multiple periodicity
  - Not easy to interpret

Our Proposal

- An alternative formulation

$$\lambda(t) = \sum_{k=0}^{p} c_k \cos(v_k t + \phi_k),$$

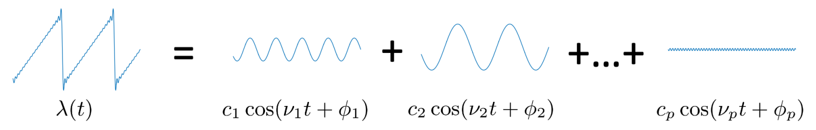
frequencies v_k, amplitudes c_k, phases ϕ_k.

- An alternative formulation

$$\lambda(t) = \sum_{k=0}^{p} c_k \cos(v_k t + \phi_k),$$

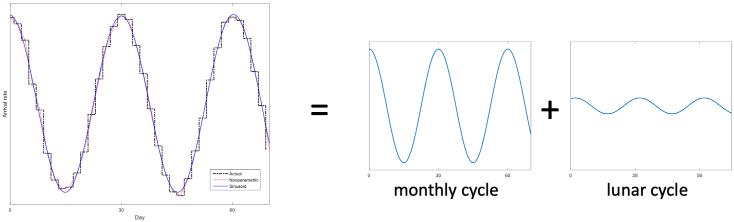
frequencies v_k, amplitudes c_k, phases ϕ_k.

- Flexibility: any periodic or non-periodic functions can be approximated (Fourier analysis)



Our Proposal

- Interpretability

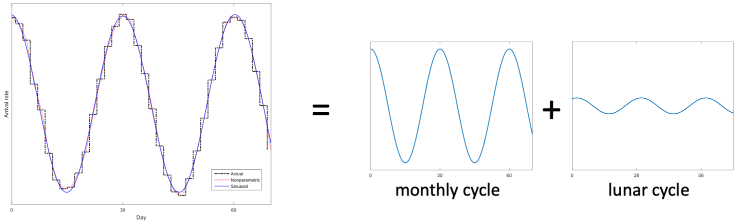


Example (Ed Kaplan) Strong monthly arrivals cycle
Nonparametric and sinusoids fit equally well

...but sinusoids also reveal hidden frequency components in arrival patterns

Our Proposal

- Interpretability



Example (Ed Kaplan) Strong monthly arrivals cycle
Nonparametric and sinusoids fit equally well

...but sinusoids also reveal hidden frequency components in arrival patterns

- May open ways to tractable analysis [Eick et al., 1993]

ESTIMATION

Estimating Arrival Rates from the Data

- Data: $t_1 < t_2 < \cdots < t_N$ are time stamps of past arrivals

Estimating Arrival Rates from the Data

- Data: $t_1 < t_2 < \cdots < t_N$ are time stamps of past arrivals
- To estimate the frequencies, use spectral (Fourier) analysis
- To estimate the amplitudes and phases, use least square estimators

Frequency Identification

- Discrete Fourier transform:

$$\tilde{N}(v) \triangleq \frac{1}{T}\left|\int_0^T e^{-2\pi i v t} dN(t)\right|$$
$$= \frac{1}{T}\left|\sum_{i=1}^{N} e^{-2\pi i v t_i}\right|$$

to approximate

$$\tilde{\lambda}(v) = \frac{1}{T}\left|\int_0^T e^{-2\pi i v t} \lambda(t) dt\right|.$$

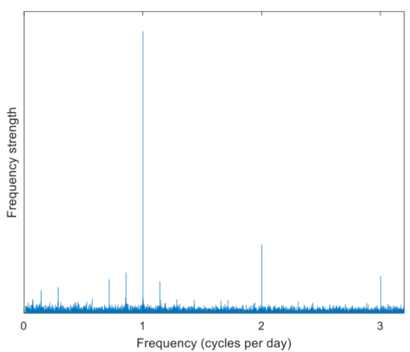
Frequency Identification

- Discrete Fourier transform:

$$\tilde{N}(v) \triangleq \frac{1}{T}\left|\int_0^T e^{-2\pi i v t} dN(t)\right|$$
$$= \frac{1}{T}\left|\sum_{i=1}^{N} e^{-2\pi i v t_i}\right|$$

to approximate

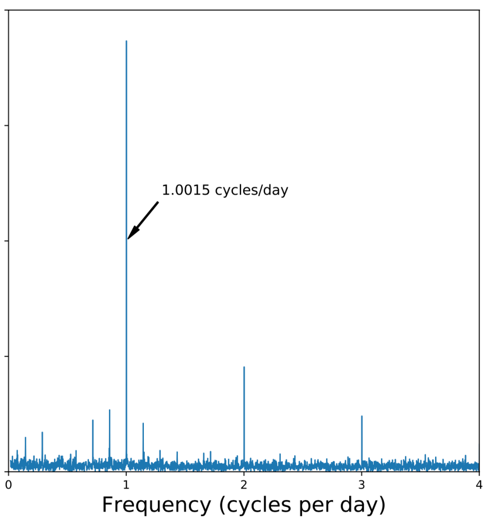
$$\tilde{\lambda}(v) = \frac{1}{T}\left|\int_0^T e^{-2\pi i v t} \lambda(t) dt\right|.$$



- Ideally we should see the right

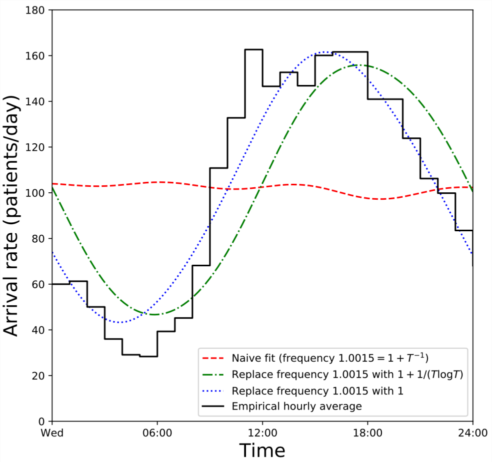
But...

- In fact, because of the noise in $N(t)$, and the leakage (finite T), we are more likely to see



Not a Big Deal? Or...

- Frequency estimation error cannot be larger than $O(1/T)$ for consistent amplitude recovery



The Solution

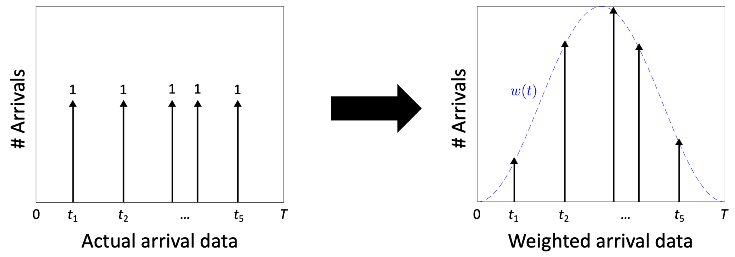
- Our innovation: Weight the number of arrivals at time t with a window function $w(t)$.

$$\tilde{N}^w(v) \triangleq \frac{1}{T}\left|\sum_{i=1}^{N} w(t_i)e^{-2\pi i v t_i}\right|$$

The Solution

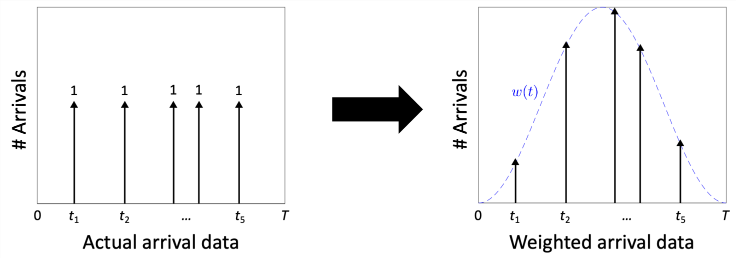
- Our innovation: Weight the number of arrivals at time t with a window function $w(t)$.

$$\tilde{N}^{w}(v) \triangleq \frac{1}{T}\left|\sum_{i=1}^{N} w(t_i)e^{-2\pi i v t_i}\right|$$



- Our innovation: Weight the number of arrivals at time t with a window function $w(t)$.

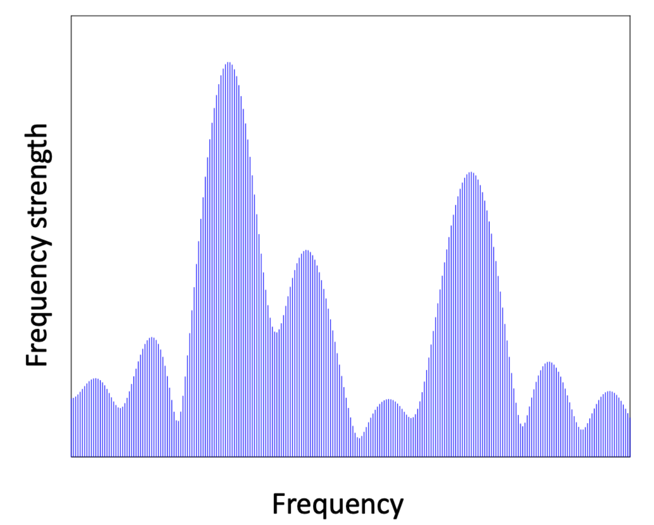
$$\tilde{N}^w(v) \triangleq \frac{1}{T}\left|\sum_{i=1}^{N} w(t_i)e^{-2\pi i v t_i}\right|$$



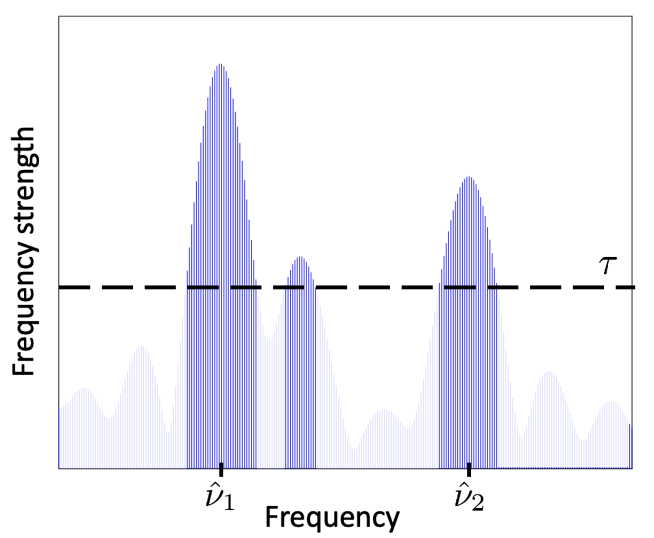
- Looks biased, but works: $\|\hat{v}_k - v_k\| = O(1/T)$ even when v_k and v_{k+1} are $O(1/T)$ close.

The Proposed Procedure

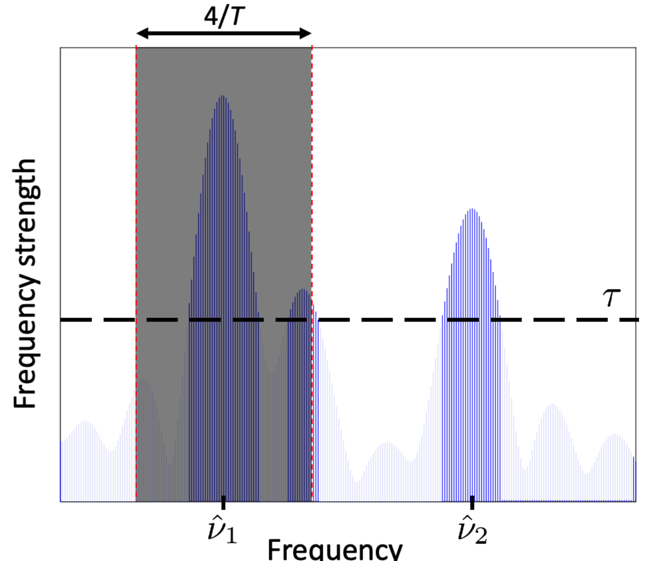
1. Compute the windowed DFT:



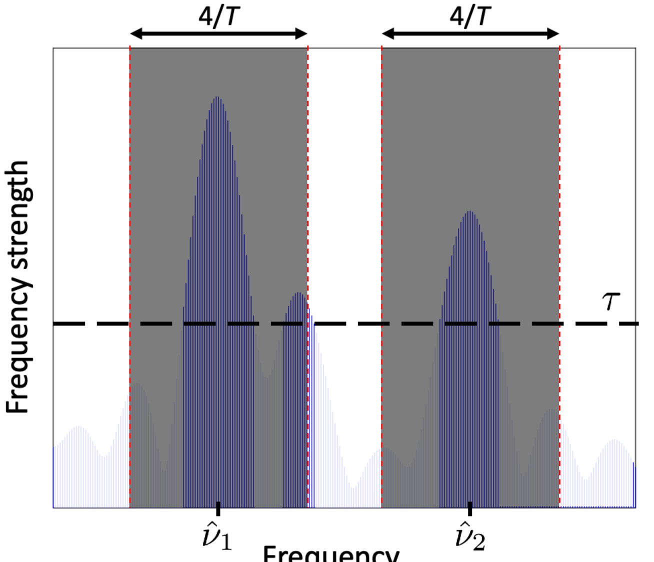
2. Compute a data-driven threshold τ:



3. Pick peaks above τ, remove a neighborhood:



4. Repeat until no peaks are above τ:



5. Based on the estimated frequencies $\hat{v}_k$, we can proceed to estimate the amplitudes and phases by the least squares.
    - We can reorganize the observations into buckets of width dt: $[0, dt]$, $[dt, 2dt], \ldots, [T - dt, T]$.
    - The observed Y is 0 or 1 for that bucket.

5. Based on the estimated frequencies $\hat{v}_k$, we can proceed to estimate the amplitudes and phases by the least squares.
    - We can reorganize the observations into buckets of width dt: $[0, dt]$, $[dt, 2dt], \ldots, [T - dt, T]$.
    - The observed Y is 0 or 1 for that bucket.
    - Least squares: find $c_{k,1}$ and $c_{k,2}$ so that

$$\hat{\lambda}(t) = \sum_{k=0}^{p} c_k \cos(\hat{v}_k t + \phi_k) = \sum_{k=0}^{p} c_{k,1} \cos(\hat{v}_k t) + c_{k,2} \sin(\hat{v}_k t)$$

    minimizes the MSE of the T/dt observations. The same as linear regression.

5. Based on the estimated frequencies $\hat{v}_k$, we can proceed to estimate the amplitudes and phases by the least squares.
   - We can reorganize the observations into buckets of width dt: $[0, dt]$, $[dt, 2dt], \ldots, [T - dt, T]$.
   - The observed Y is 0 or 1 for that bucket.
   - Least squares: find $c_{k,1}$ and $c_{k,2}$ so that

$$\hat{\lambda}(t) = \sum_{k=0}^{p} c_k \cos(\hat{v}_k t + \phi_k) = \sum_{k=0}^{p} c_{k,1} \cos(\hat{v}_k t) + c_{k,2} \sin(\hat{v}_k t)$$

     minimizes the MSE of the T/dt observations. The same as linear regression.
   - If $dt \to 0$, then $(X^T X)^{-1} X^T Y$ has a closed form.

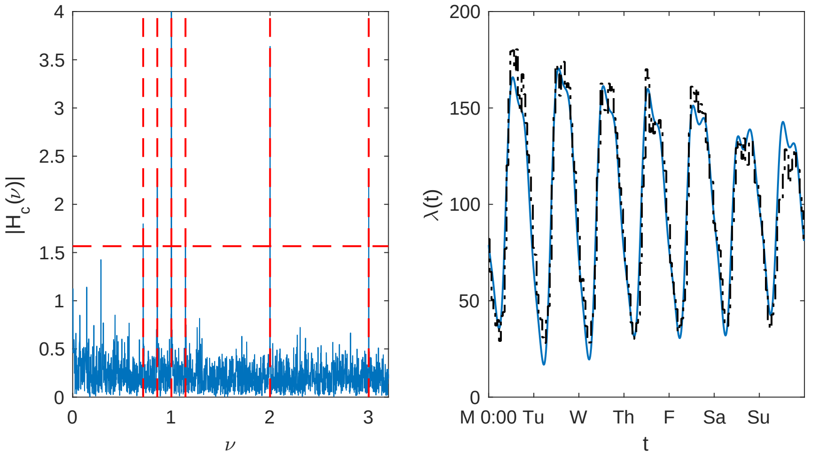
Empirical Study

Arrival Data from an Emergency Department

Data characteristics:

- Time stamps of 168,392 patent arrivals from 2014 Jan to 2015 Sept ($T = 652$ days)
- Emergency Severity Index (ESI) level of each patient
  - Level 1 most severe (e.g., cardiac disease); level 5 least severe (e.g., rash)
- We analyze ESI level 2 and level 3 to 5 separately (level 2 are treated in a separate ward)

- 66,240 patient arrivals
- Estimated frequencies: $\hat{\nu}_1 = 1.00$, $\hat{\nu}_2 = 2.00$, $\hat{\nu}_3 = 3.00$, $\hat{\nu}_4 = 0.714$, $\hat{\nu}_5 = 0.857$, $\hat{\nu}_6 = 1.143$



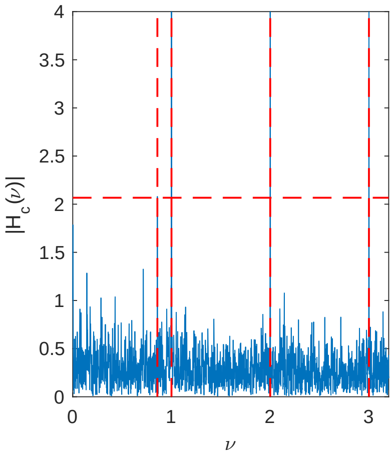
- $\hat{v}_1 = 1.00$, $\hat{v}_2 = 2.00$, $\hat{v}_3 = 3.00$ make up the daily cycle.
- $\hat{v}_4 = 0.714$ (5/7), $\hat{v}_5 = 0.857$ (6/7), $\hat{v}_6 = 1.143$ (8/7) make up the weekly cycle.

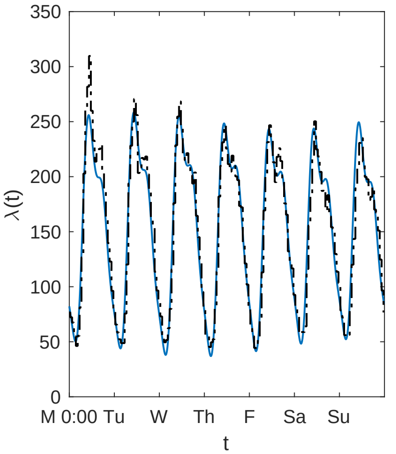
Interpretation

- $\hat{v}_1 = 1.00$, $\hat{v}_2 = 2.00$, $\hat{v}_3 = 3.00$ make up the daily cycle.
- $\hat{v}_4 = 0.714$ (5/7), $\hat{v}_5 = 0.857$ (6/7), $\hat{v}_6 = 1.143$ (8/7) make up the weekly cycle.
- There are two peaks in a day; the intensity of arrivals fade steadily into the weekend.

ESI Level 3 to 5

- 99,205 patient arrivals
- Estimated frequencies: $\hat{v}_1 = 1.00$, $\hat{v}_2 = 2.00$, $\hat{v}_3 = 3.00$, $\hat{v}_4 = 0.857$





Interpretation

- Only one weekly cycle is present $\hat{v}_4 = 0.857$. The weekly cycle is weaker than level 2.

Interpretation

- Only one weekly cycle is present $\hat{v}_4 = 0.857$. The weekly cycle is weaker than level 2.
- Unable to capture the localized spikes on Monday, will need more weekly cycles

Interpretation

- Only one weekly cycle is present $\hat{v}_4 = 0.857$. The weekly cycle is weaker than level 2.
- Unable to capture the localized spikes on Monday, will need more weekly cycles
- In both cases
    - No monthly cycles are identified
    - No seasonal cycles are identified, probably because T is not large enough

Summary

We propose a sine-wave-based approach to the modeling and estimation of non-stationary arrival processes. Compared to the common approach:

- Not requiring prior knowledge of periods
- Can handle conflated multiple periodicity
- Much sparser ($3p$ vs. hundreds of parameters)
- May provide interpretable insights

- Computation is not straightforward
- Sensitive to the threshold
- May miss localized spikes