Memory in Bursty Systems

Fangjian Guo
(Final-year undergrad in computer science)
richardkwo@gmail.com
http://richardkwo.net
In collaboration with Prof. Tao Zhou and Ms. Zimo Yang

Web Sciences Center, UESTC

Jan 23, 2013

Inter-event time series in complex systems

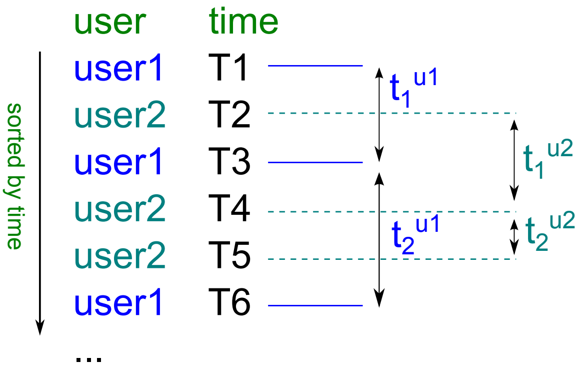
Complex systems usually refer to those made up of a large number of components interacting in a complex structure.

- Traffic, earthquakes, climate, and even the society are examples of complex systems.
- The online human activities are also a good data source for research
- *"emergence"*, i.e. the system as a whole exhibit some properties that are not obvious from its individual components.

Inter-event time series in complex systems

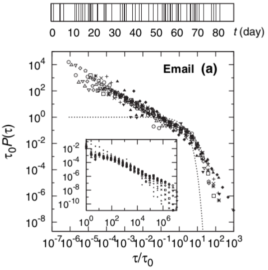
Inter-event time series is time series corresponding to the time elapsed during *two consecutive events*.

- Here we focus on time series on the *individual level*.



Statistical properties

Notable statistical properties are found for inter-event time series in complex systems.



- **Burstiness**: *"short timeframes of intense activity followed by long times of no or reduced activity".*
- **(positive) Memory**: *long intervals between activities tend to be followed by long intervals and short followed by short.*

Characterization

Burstiness can be measured with the moments of $\{t_i\}$ (other measures are possible).[1]

$$B = \frac{\sigma - m}{\sigma + m}$$

Memory, as a temporal correlation, can be measured with *autocorrelation function*

$$M = \frac{1}{n-1} \sum_{i=1}^{n-1} \frac{(T_i - m_1)(T_{i+1} - m2)}{\sigma_1 \sigma 2}$$

- Here, we use *lag-1 (1st order) autocorrelation*.
- It is simply the Pearson's autocorrelation between $\{t_1, t_2, \cdots, t_{n-1}\}$ and its lag-1 counterpart $\{t_2, t_3, \cdots, t_n\}$.

[1] Goh, K-I., and A-L. Barabási. "Burstiness and memory in complex systems." EPL (Europhysics Letters) 81.4 (2008): 48002.

Mechanisms

What caused the *burstiness* and *memory*?

- Marginal distribution: **Power-law**
- **Ordering**: the structure of interdependence among $\{t_i\}$

Are these two properties **orthogonal** (as suggested by Barabási et al.) or **inter-related**?

- We already know that, *burstiness* is partially attributable to *power-law*.
- Then, does *power-law* have any implication on *memory*?

Modeling correlated power-law series

We capture the two mechanisms (*marginal* and *interdependence*) with the following two-step process of generating power-law series.

1. **Independently** sample n elements from the sample power-law marginal $P(t) \sim t^{-\alpha}$.
2. Change the interdependence among elements by **shuffling** the ordering while preserving their values, one arbitrary ordering results in one $\{t_1, t_2, \cdots, t_n\}$.

Such a family of series,

- Follow a power-law.
- Cover a wide space of interdependence among elements (state space of $n!$).
  For example, $\{t_i\}$ sorted in increasing order would have stronger memory effect than one in random order.

Min/Max sort and the bounds for memory

To be formal, consider the series $\{t_1, t_2, \cdots, t_n\}$, each sampled independently from the same distribution $F(x)$. Then, we sort them in increasing order as $\{t_{(1)}, t_{(2)}, \cdots, t_{(n)}\}$, such that

$$t_{(1)} \leq t_{(2)} \leq \cdots \leq t_{(n)}.$$

Now we apply a permutation $\theta : \{1, 2, \cdots, n\} \rightarrow \{1, 2, \cdots, n\}$ to the originals series, resulting in a new series $\{t_{\theta(1)}, t_{\theta(2)}, \cdots, t_{\theta(n)}\}$. Among the possible permutations (state space of size $n!$), there would be **one series with maximum memory** and **one with minimum memory**[2].

$$\{t_{(i)}\} \xrightarrow{\theta_{\max}} M_{\max}$$

$$\{t_{(i)}\} \xrightarrow{\theta_{\min}} M_{\min}$$

[2]Hallin, Marc, Guy Melard, and Xavier Milhaud. "Permutational extreme values of autocorrelation coefficients and a Pitman test against serial dependence." The Annals of Statistics 20.1 (1992): 523-534.

Min/Max sort and the bounds for memory

$\theta_{\max}$ obtains the maximum memory by first arranging the odd elements of order statistics in increasing order, followed by even elements in decreasing order, which is

$$\begin{aligned} & t_{(1)}, t_{(3)}, \cdots, t_{(2l-1)}, t_{(2l)}, t_{(2l-2)}, \cdots, t_{(4)}, t_{(2)} \quad (n = 2l), \\ & t_{(1)}, t_{(3)}, \cdots, t_{(2l-1)}, t_{(2l+1)}, t_{(2l)}, \cdots, t_{(4)}, t_{(2)} \quad (n = 2l+1). \end{aligned} \tag{1}$$

On the contrary, quite interestingly, $\theta_{\min}$ arranges the order statistics by interlacing the even and odd terms when $n = 2l$ or small and big terms when $n = 2l + 1$, i.e.

$$\begin{aligned} & t_{(2l)}, t_{(1)}, t_{(2l-2)}, \cdots, t_{(2l-3)}, t_{(2)}, t_{(2l-1)} \quad (n = 2l), \\ & t_{(2l)}, t_{(2)}, \cdots, t_{(l)}, \cdots, t_{(1)}, t_{(2l+1)} \quad (n = 2l+1). \end{aligned} \tag{2}$$

Deriving the bounds of memory for shuffled power-law

The theoretical bounds for M in the sense of

- expected value w.r.t the sampling process
- $n \to \infty$

$$M_{\max} = \lim_{n\to\infty} E\left[\max_{\theta} M(\{t_{\theta(i)}\})\right], M_{\min} = \lim_{n\to\infty} E\left[\min_{\theta} M(\{t_{\theta(i)}\})\right]$$

Methods

- $\alpha > 3$ (convergent 2nd moment): Probabilistic method (order statistic distribution $\to$ expected value $\to$ limit in $n \to \infty$)
- $\alpha \leq 3$ (diverging moments): Approximating random samples with equiprobable slices.

Probabilistic method ($\alpha > 3$)

$$M = \frac{1}{n-1}\sum_{i=1}^{n-1}\frac{(t_i - m_1)(t_{i+1} - m2)}{\sigma_1\sigma_2}, \tag{3}$$

A little algebra,

$$E[M] = \frac{1}{\sigma(\alpha)^2}(\frac{1}{n-1}\sum_{i=1}^{n-1}E[t_i t_{i+1}] - m(\alpha)^2). \tag{4}$$

Sample moments are replaced by population moments,

$$m(\alpha) = \frac{\alpha-1}{\alpha-2}, \sigma(\alpha)^2 = \frac{\alpha-1}{\alpha-3} - m(\alpha)^2$$

Probabilistic method ($\alpha > 3$)

Upper bound

$$M_{\max} = \lim_{n\to\infty} E[M_{\theta_{\max}}] = \frac{1}{\sigma(\alpha)^2}(\lim_{n\to\infty}\frac{1}{n-1}E[S_{\theta_{\max}}] - m(\alpha)^2)$$

Recalling $\theta_{\max}$,

$$\frac{1}{n-1}E[S_{\theta_{\max}}] = \frac{1}{2l-1}\sum_{i=1}^{2l-2}E[t_{(i)}t_{(i+2)}] + \frac{1}{2l-1}E[t_{(2l)}t_{(2l-1)}],$$

To obtain the expected value of bivariate product, we adopt the bivariate distribution of **order statistics**,

$$f_{t_{(j)},t_{(k)}}(x,y) = n!\frac{[F(x)]^{j-1}}{(j-1)!}\frac{[F(y)-F(x)]^{k-1-j}}{(k-1-j)!}\frac{[1-F(y)]^{n-k}}{(n-k)!}f(x)f(y) \quad (x \le y)$$

Probabilistic method ($\alpha > 3$)

$$
\begin{aligned}
\mathrm{E}[t_{(i)}t_{(i+2)}] &= \iint_{1\le x\le y<\infty} xyf_{t_{(i)},t_{(i+2)}}(x,y)\,dxdy \\
&= \frac{\Gamma(2l+1)}{\Gamma(2l+1-2c)}\frac{\Gamma(2l-i+1-2c)}{\Gamma(2l-i-1)}\frac{1}{[(2l-i)-c][(2l-i)-\alpha c]} \qquad (1
\end{aligned}
$$

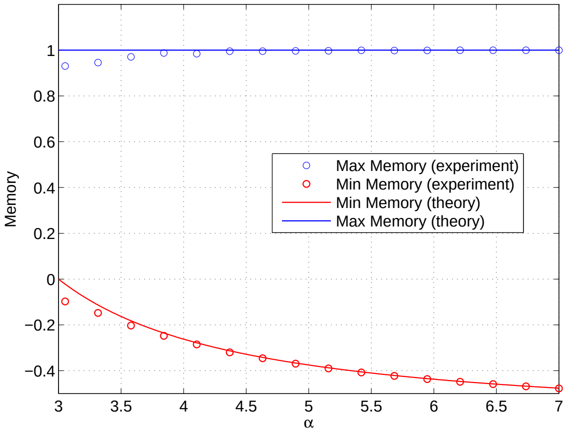
Aha, add them up and get the limit,

$$M_{\max} = \lim_{n\to\infty}\mathrm{E}[M_{\theta_{\max}}] = 1 \quad (\alpha > 3)$$

Similarly, the lower bound is obtained as

$$M_{\min} = \lim_{n\to\infty}\mathrm{E}[M_{\theta_{\min}}] = \frac{1}{\sigma(\alpha)^2}\left[2B\left(\frac{1}{2};\frac{1}{m(\alpha)},\frac{1}{m(\alpha)}\right) - m(\alpha)^2\right] \quad (\alpha > 3)$$

Probabilistic method ($\alpha > 3$)



Equiprobable slices approximation ($\alpha \leq 3$)

When $\alpha \leq 3$,

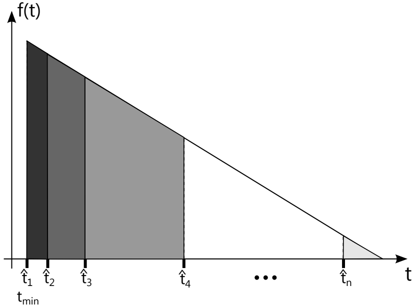
- Equating sample moments with population moments is no longer invalid
- The distribution of M itself is intractable

Our approximation solution:

- substituting the random samples with determinant values in the limit of large n

Equiprobable slices approximation ($\alpha \leq 3$)

Cutting the area under $f(x)$ into equiprobable slices



$$M = \frac{1}{\sigma^2}\left(\frac{1}{n-1}\sum_{i=1}^{n-1} t_i t_{i+1} - m^2\right)$$
$$= \frac{\frac{1}{n-1}\sum_{i=1}^{n-1} t_i t_{i+1} - m^2}{\frac{1}{n}\sum_{i=1}^{n} t_i^2 - m^2},$$

The asymptotic behavior of diverging statistics

By using the approximation scheme, we can know "how fast" the statistics/moments diverge — *the asymptotic behavior of diverging statistics.*

e.g.

$$\begin{aligned} \hat{s}_{\theta_{\max}} &= \frac{1}{n-1}\left(\sum_{i=1}^{2l-2} \hat{t}_i \hat{t}_{i+2} + \hat{t}_{2l}\hat{t}_{2l-1}\right) \quad (n=2l) \\ &= \frac{1}{n-1} n^{2c}\left[\sum_{k=2}^{n-1}(k^2-1)^{-c} + 2^{-c}\right] \quad (\alpha<3), \end{aligned} \tag{5}$$

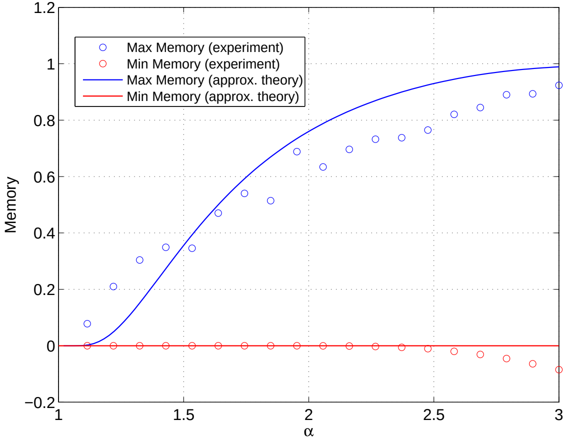
$$\overline{\hat{t}^2}_{\theta_{\max}} = \frac{1}{n}\sum_{i=1}^{n} \hat{t}_i^2 = n^{2c-1}\sum_{k=0}^{n-1}(k+1)^{-2c} \quad (\alpha<3). \tag{6}$$

Equiprobable slices approximation ($\alpha \leq 3$)

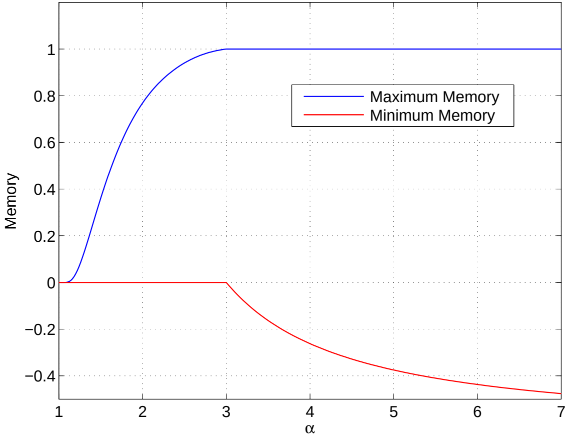
$$
\begin{aligned}
M_{\max} = \lim_{n\to\infty} \mathrm{E}[M_{\theta_{\max}}] &\approx \lim_{n\to\infty} \hat{M}_{\theta_{\max}} \\
&\approx \lim_{n\to\infty} \frac{\sum_{k=2}^{n-1}(k^2-1)^{-c} + 2^{-c}}{\sum_{k=0}^{n-1}(k+1)^{-2c}} \qquad (1 < \alpha \leq 3)
\end{aligned}
$$

$$
M_{\min} = \lim_{n\to\infty} \mathrm{E}[M_{\theta_{\min}}] \approx \lim_{n\to\infty} \hat{M}_{\theta_{\min}} = 0
$$

Equiprobable slices approximation ($\alpha \le 3$)



Theoretical bounds



Empirical results compared to theory

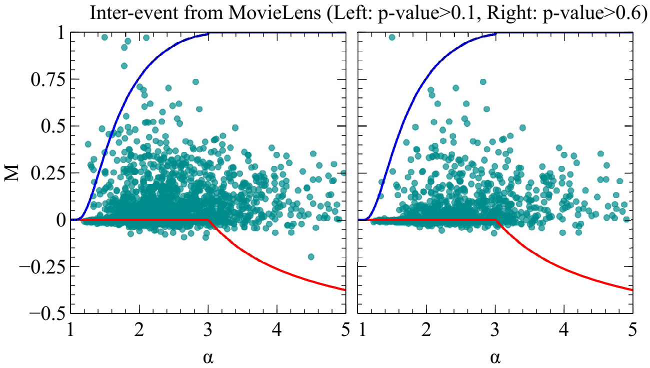
Getting empirical power-law inter-event time series

1. Online human activities: scoring, visiting, tweeting, etc.
2. Event time $\rightarrow$ Inter-event time (1 series for 1 user)
3. Choose those long enough ($n > 200$)
4. Fitting to power-law[3] and choose $p - value > 0.1$.

[3] A. Clauset, C. Shalizi, and M. Newman. Power-law distributions in empirical data. SIAM Review, 51(4):661–703, 2009.

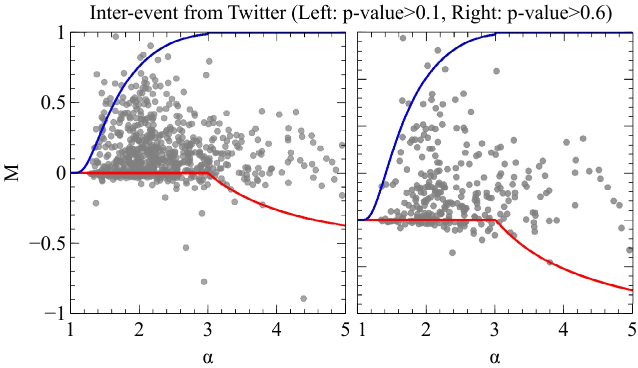
Empirical results compared to theory

MovieLens



Empirical results compared to theory

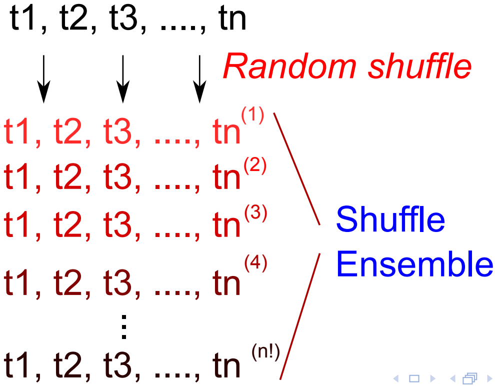
Twitter



Towards a better measure of memory

1. How to measure memory with **higher resolution**? (Is $M = 0.001$ really weak?)
2. How to measure memory as a dimension of “ordering”?

Our solution: comparing a series’ memory with its **shuffle ensemble**



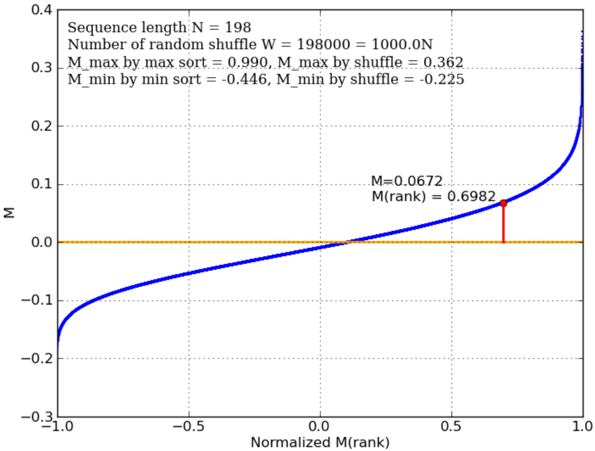
Towards a better measure of memory

Method:
For series $\mathbf{T} = \{t_1, t_2, \cdots, t_n\}$,

1. Random-shuffle the original series to get W series, which forms a subset of shuffle ensemble $\{\mathbf{T}^{(1)}, \mathbf{T}^{(2)}, \cdots, \mathbf{T}^{(W)}\}$
2. Compute memory for both the original and shuffled counterparts: M and $\{M^{(1)}, M^{(2)}, \cdots, M^{(W)}\}$
3. Get the **ranking** of M among $\{M^{(1)}, M^{(2)}, \cdots, M^{(W)}\}$ as R
4. Linearly mapping the ranking R to $[-1, +1]$ as the new memory

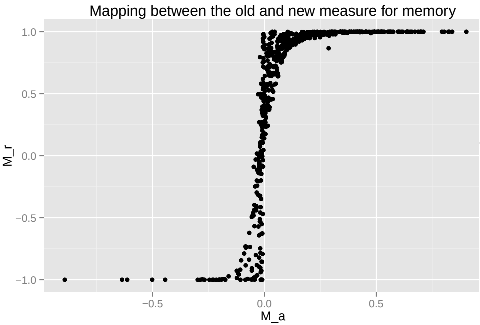
$$M^{(R)}(\mathbf{T}; W) = \frac{2}{W}(R-1) - 1$$

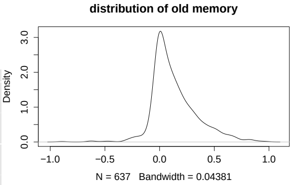
Towards a better measure of memory

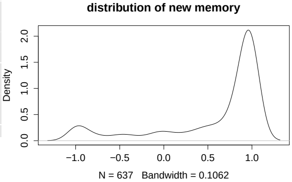


Towards a better measure of memory

"Mapping" between the old and new M measure (from **Twitter** data)







Thank you

Thanks a lot, any question?