MSc Thesis

Modelling of communication dynamics

Szabolcs Vajna

Supervisor: **Prof. János Kertész**
Department of Theoretical Physics
Institute of Physics

Budapest University of Technology and Economics
2012

Contents

1 Introduction

Computational social science [1] is a rapidly developing area of scientific research. While people are doing their usual daily activities – for example they are using their mobile phones, their credit cards, etc. – they leave digital footprints behind themselves. These footprints can be used to make important observations and measurements about the human behaviour, which was impossible a few years ago. The two main directions of computational social science are the analysis of social networks and *human dynamics*. Beside academic interest, studying human dynamics may have practical usage as well. One example is analysing human mobility that is essential in modelling (and anticipating) the spread of diseases. Social and economic phenomena are driven by individual human actions, thus timing of human activities is of central interest. This research is truly multidisciplinary and physicists play an important role in it. The main aim of physics is to give mathematical description about the phenomena in Nature. Social phenomena can also be considered as part of Nature and empirical studies, and even experiments can be made on the digital datasets. Physicists are used to making simple (and more complex) models that are compared to the data every time. Nowadays more and more datasets are becoming available for scientific research [2, 3, 4, 5], and physicists can analyse these datasets similarly to those obtained from the physical experiments. The essential difference is that in social sciences we usually cannot influence the systems or prepare subsystems. The only thing we can do is to watch their intrinsic dynamics. (From this point of view human dynamics is more an observational science like astronomy.) Due to fluctuations and external interventions, like rare phenomena of non human-related Nature that has effects on the human related world, we are able to analyse subsystems and the response of the systems to disturbances. The stochastic character, the interaction of many constituents and the data deluge have made physicists interested in the field of human dynamics and network theory and they have had a major impact on these fields.

1.1 Previous results from literature

A large number of detailed electronic datasets have become available for scientific research recently, which contain for example the timing of email [2, 3] and telecommunication [4] actions. In human dynamics the most important and most studied quantity is the *interevent time*, that is the time elapsed between two consecutive actions of the same kind of activity. For many types of activities the interevent time distribution shows power-law decay. This means that very active and long inactive periods can be observed on wide range of scales. The timing of events is not homogeneous, bursts of frequently repeated actions are separated by long inactive periods. We are going to refer to this type of activity pattern as *bursty behaviour*. This is in contrast with the previously assumed Poissonian human behaviour where long interevent times are exponentially rare. The bursty behaviour can be characterised by the exponent of decay of the interevent time distribution. The interevent time distribution is power-tailed for example for the library loans [6], for email communication [2, 6] and for telecommunication patterns [7]. The interevent time distributions for the surface mail based communication of Darwin, Einstein and Freud are also power-tailed [8].

The first model to describe bursty behaviour was introduced by Albert-László Barabási [9]. His publication in 2005 triggered much activity in the study of human dynamics. He wanted to model the distribution of *waiting times* between reading and replying to an email that might be responsible for the heavy tail of the *interevent time* distribution in the email communication. In this

model people are assumed to have a list of size L that contains their tasks. Every task has a priority x_i ($i = 1 \dots L$) from a $\rho(x)$ distribution. At each time step with probability p the user chooses the task with maximum priority, and with probability $1-p$ he chooses a task uniformly. The chosen task is then executed, and a new task comes with priority chosen from the distribution $\rho(x)$. It can be shown that the waiting time distribution in the model has a power-law decay $P_{wt}(t) \sim t^{-\beta}$ with exponent $\beta = 1$ [9, 10]. This model is in agreement with the empirical results: both of the waiting time and the interevent time distribution can be approximated by a power decay with exponent 1. The finite length of the list models the finite memory of the users. We should note that if the length of the list is not fixed then the decay of the interevent time distribution changes. If we assume that tasks are coming and executed with the same rate then the exponent for the waiting time is $\beta = 3/2$. This is the exponent that has been observed in Darwin's, Einstein's and Freud's surface mail patterns [8]. In these cases the length of the list is given by the quantity of the letters in the drawer, and the quantity can fluctuate in wide range. This can cause the difference between the exponents in the surface mail and the email communication.

Vázquez in his 2006 paper argued about the discrete universality classes in human dynamics [6]. For example web-browsing, library loans and email communication are in the $\beta = 1$ universality class, while the surface mail communication is in the $\beta = 3/2$ universality class. In contrast there is increasing evidence against the existence of universality classes. Zhou *et al* studied Netflix users sorted by activity [5] (Netflix is a webpage on which movies can be rated by the users). They found that the interevent time distribution is different for the groups with different activity, and the exponent shows a monotonic dependence on the activity of the group. Similar results have been reported in the analysis of short message communication [11]. The Barabási-model has a generalisation in which the power-law exponent is tunable. That model was introduced in the same paper as the original one, and its aim was just to simplify the calculations, but it turned out to have some relevance. In this modified model the probability of choosing an element is a power function of its priority: $\Pi(x) \sim x^s$ and the power-law exponent is $1 + 1/s$. The $s \to \infty$ limit of the modified model gives the $p \to 1$ limit of the original model.

In 2008 Malmgren *et al* proposed a totally different explanation for the heavy tails in the email communication [12]. They claimed that the presence of heavy tails in the interevent time distribution is simply the consequence of the circadian and weekly cycles of human activity. They proposed a model, the cascading nonhomogeneous Poisson process, which takes into account both the periodic patterns of activity and the individual's tendency to continue participating in an activity. The primary process is a nonhomogeneous Poisson process with a rate function with weekly periodicity, and every event generated by this process starts a second Poisson process with higher (but constant) rate. The distribution of the length of the second process also has to be fitted to the data. They also calculated the Fano and Allen factors to decide if there is a deeper correlation between the events, and they stated that the empirical email data is not more correlated (bursty) than a Poisson process [13].

In contrast to this Hang-Hyun Jo *et al* at Aalto University introduced a deseasoning method [14] that can remove the circadian and weekly patterns from the time series of mobile communication events for individuals. They found that the heavy tail of the interevent time distribution remains robust with respect to the deseasoning procedure. This indicates that the human task execution-based mechanism is a possible cause of the remaining burstiness in temporal mobile phone communication patterns.

Wu *et al* studied a model where a Poisson process initiates a task-queuing model [15]. They

also included interaction between the lists of the users, and with this model they could explain the bimodal shape of the interevent time distribution seen in the short-message communication. By bimodal distribution they mean that up to some time t^* the distribution follows a power-law, and after that it changes to exponential (and not to a product of a power and exponential function).

There are several more models of human dynamics beyond the task queuing models and the cascading nonhomogeneous Poisson process. One interesting attempt is the interest-driven model [16], in which the frequency of events is determined by the interest which is influenced by the occurrence of events.

The origin of the observed burstiness is a controversial issue. Some scientists claim that the cascading Poisson process gives a sufficient explanation, while others argue that in order to interpret all observations intrinsic correlations have to be assumed.

Zhou's *"Towards the Understanding of Human Dynamics"* review article [17] is recommended to read to get a deeper insight to the empirical observations on different activities with power decaying interevent time distributions and to some models (Barabási- and the interest-driven models).

1.2 Dataset and analysed quantities

I do my research in collaboration with the research group at Aalto University in Finland. This research group has been given access to the time stamped records of mobile phone calls (MPC) and short messages (SM) from January 1, 2007 over a period of 120 days by an unnamed European operator whose national market share is about 20%. The dataset is anonymous and it contains the directions of the calls and the starting and ending times with resolution of seconds. The MPC dataset contains $N = 5.2 \times 10^6$ users, $L = 10.6 \times 10^6$ links and $C = 322 \times 10^6$ calls, and the SM dataset contains $L = 8.5 \times 10^6$ links and $C = 114 \times 10^6$ messages.

The research group has achieved many results in analysing the network properties constructed from the MPC dataset [18, 19]. They have also started to analyse the time series in connection with human dynamics. Now I am introducing the most important quantities that were computed from the dataset.

Interevent time distribution

The most studied quantity is the interevent time distribution. The interevent time is defined as the time elapsed between the *end* of the last call and the *beginning* of the next, i.e. it measures the length of inactive periods. Let us focus on a single user. Let t_i denote the time when the i^{th} call is started by the user and d_i denote the duration of the i^{th} call. Then the i^{th} interevent time is defined by $\tau_i = t_{i+1} - (t_i - d_i)$. For the probability density or mass function we are going to use the $P_{ie}(t)$ notation. This function measures the inhomogeneity of the calls but states nothing about the correlations between the calls.

Autocorrelation function

One way to measure correlations in time series is to calculate the *autocorrelation function* of the *event sequence*. The timing of calls can take only integer values for all digital databases. We define the contracted event sequence $\{\tilde{t}_i\}$ in which the duration of calls are set to 0: $\tilde{t}_i = t_i - \sum_{j=1}^{i-1} d_i$. Let X be the indicator function of starting a call: $X(t) = 1$ if $t = \tilde{t}_i$ for some i and 0 otherwise. The

autocorrelation function for this contracted event sequence is the following:

$$A(t) = \frac{\langle X(t+\tau)X(\tau)\rangle_\tau}{\langle X(\tau)\rangle_\tau} \tag{1}$$

In the denominator we used the property that $X^2 = X$ for indicator functions. Contracted event sequence is used because we would like to remove the repulsion of calls due to their finite duration.

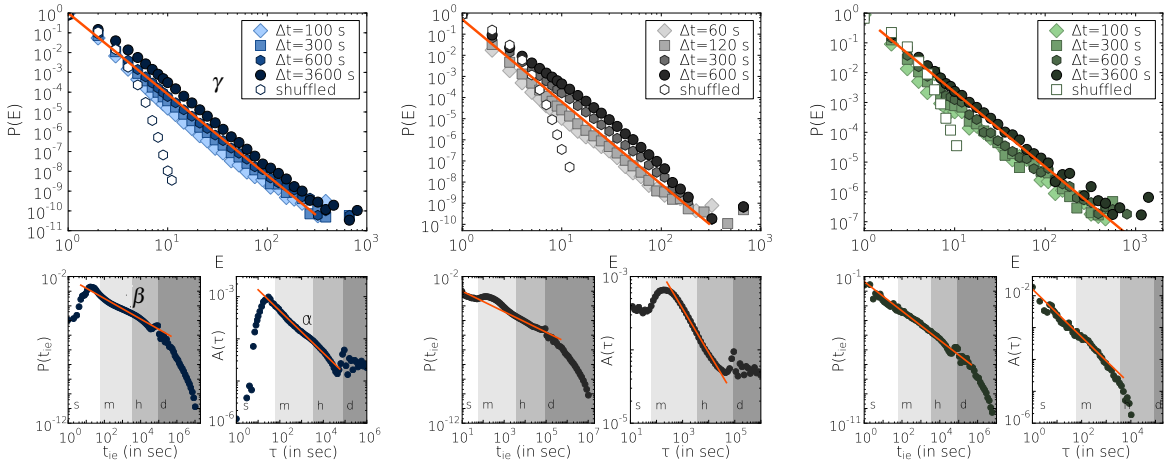
Number of bursty events

The autocorrelation function is a two-point function. The research group was interested in deeper correlations and defined a new measure, the distribution of the number E of bursty events in a bursty period. Bursty periods are defined by a time window Δt: two consecutive events are in the same bursty period if the time between them is smaller than the time window, i.e. $t_{i+1} - (t_i + d_i) \leq \Delta t$. It is easy to show that the distribution of the number of bursty events $P_{E,\Delta t}(n) = \mathbb{P}(E_{\Delta t} = n)$ is exponentially decaying in n for any processes with independent interevent times.

$$\mathbb{P}_{E,\Delta t}(n) = \left(\int_0^{\Delta t} P_{ie}(t)dt\right)^{n-1}\left(1 - \int_0^{\Delta t} P_{ie}(t)dt\right) \tag{2}$$

1.3 Previous results of the research group

The three characteristic functions defined in the previous section are similar in different communication channels (fig.1, mobile phone call, short message and email communication). This signalises that there is some deep process that organizes human communication dynamics and gives hope to find a model that can capture the essential properties of this process. The time series are



(a) Mobile phone call (b) Short Message (c) Email

Figure 1: Márton Karsai's measurements on different communication databases [20]. The upper panels in the plots show the P_E distribution in my notation. The exponents referring to the scaling regions can be found in tab.1.

very rich in phenomena and we are looking for the most characteristic properties as the beginning

of our analysis. This is why the duration of calls are extracted from the data. The effect of very short and long calls or the interaction with incoming calls may be studied in the future. In the case of email and short-message communication the effect of interaction is expected to be much higher than in the mobile call communication, because in the latter the two participants can react to each other at the same call.

We can observe two time scales in the autocorrelation function and in the interevent time distribution in fig.1, they appear as a cutoff in the functions. The lower cutoff starts at ~ 20 seconds, which is the consequence of the typical reaction time after a call. The upper cutoff starts between 12 hours and a day. The cutoff can be described by an exponential or a power-law function with a very large exponent. Between these two cutoffs both of the interevent time distribution and the autocorrelation function show approximately power-law decay: $P_{ie}(t) \sim t^{-\beta}$ and $A(t) \sim t^{-\alpha}$. The observed exponents can be found in tab.1. Márton Karsai's measurements show that for the MPC

	α	β	γ
MPC	0.5	0.7	4.7
SMS	0.6	0.7	3.9
E-mail	0.75	1.0	2.75

Table 1: Exponents for the autocorrelation function (α), for the interevent time distribution (β) and for the number of bursty events distributions (γ) [20].

dataset the number of bursty events distribution has a power-law regime and the exponent ($\gamma \approx 4.1$) does not depend on the time window when it is chosen from the interval between 32 seconds and 9 hours.

The functions in fig.1 show an average over the whole dataset. These measurements were repeated with averaging just over the users within the same group of activity (total number of calls started), and similar scaling was found but with different exponents for the different groups (see supplementary information in [20]).

As it has been mentioned before, the scaling region of the characteristic functions are robust to the procedure that removes the daily and weekly patterns [14].

1.4 Questions and my results

My final aim is to find an appropriate model for human communication dynamics. I am interested in analysing the dataset at the level of individuals. Though Márton Karsai has made measurements for groups decomposed by overall activity, these groups still cover a wide range of activity. On the other hand two users with the same activity can behave very differently. I have access to a small subset of the MPC dataset (sMPC) and one part of my work is to analyse the characteristic functions on the level of individuals. I am also interested in the question whether the cascading Poisson process can make a good fit to the data or we can see some additional effect of burstiness.

These questions are discussed in section 2 and 3. In this dissertation I will show that the behaviour of individuals varies within the sMPC dataset even for individuals with similar total activity. The circadian pattern is well defined for almost every user but the weekly pattern is present only for about half of them. The long time regime of the autocorrelation function can be reproduced by any process with a proper rate function. A non negligible proportion of the users have changed their average activity during the entire period. This should be taken into account if someone wants

to use a nonhomogeneous Poisson process (which is usually defined by a weekly rate function). The power-law regime of the autocorrelation function differs from the nonhomogeneous Poisson process, and an additional Poisson cascade usually cannot fit the autocorrelation function on the entire domain.

I have also studied a family of task-queuing models (section 4) that can be thought of as a generalisation of the Barabási-model. These models have many advantages compared to the Barabási-model. In these models we can define the interevent times (not only the waiting times), and as a consequence the autocorrelation function can also be defined. I have analysed these models by numerical and analytical methods. I have shown that these models can produce power decay in the interevent time distribution and in the autocorrelation function with tunable exponents. I have proved a scaling law that connects the exponents of the autocorrelation function and the interevent time distribution in these models. This result can be extended to all renewal processes with slowly decaying interevent times.

In my thesis I cannot give answers to all questions that have arisen, and the connection between the empirical measurements and the task queuing model is also weak. However, the results of the task queuing model are more general and could be used in a wide range of problems where bursty behaviour is present (e.g. in timing of earthquakes [21], or in neuron firing sequences [22]). Finally the composition of task queuing models with a rate function is also promising.

2 Empirical measurements

I have made measurements on a small extract from the MPC dataset, hereafter referred to as sMPC. I have calculated the three characteristic functions ($P_{ie}(t)$, $A(t)$ and $P_E(n)$) for the individuals to see how much and what kind of information is lost during the averaging. I have also analysed the circadian and weekly patterns and some other properties.

2.1 The sMPC dataset

The sMPC dataset contains 70 people of 7 strength groups, where the groups are defined by the overall user activity (tab.2). The dataset covers 180 days. We refer to the users by an ID number:

Str group	1	2	3	4	5	6	7
# of calls	$0-104$	$105-208$	$209-416$	$417-834$	$835-1670$	$1671-3342$	$3343-5212$

Table 2: Number of calls for the different strength groups

the first digit refers to the activity of the user (in which strength group is the user), and the second digit distinguishes the users inside the groups (it goes from 0 to 9).

This dataset does not include the timing of the calls, it only includes the interevent time sequences. The interevent time is the time elapsed between the *end* of the last call and the *beginning* of the next. Hence the timing of calls cannot be reproduced from this dataset, the length of calls would be needed for it. This will cause problems in determining the circadian pattern and the autocorrelation function.

In the case of users with high activity there is enough data to get smooth characteristic curves. For users with lower activity these functions start to be noisy and can be sampled on fewer points. Averaging, if it is permitted, can smoothen these curves but the temporal patterns of different individuals even from the same group seem to be sometimes so different that it is hard to assume that they are samples from the same distribution. This sMPC dataset is too small to get good statistics from it, but it can show the differences and similarities between the individuals. By analysing this dataset we can find interesting properties that *should be measured on the MPC dataset afterwards*.

2.2 Interevent time distribution

The dataset contains the interevent time sequence τ_i $i = 1,2,\ldots$ for all users. The empirical interevent time distribution $P_{ie}(t)$ is the histogram of the interevent times. The interevent times take values from broad range (heavy tailed distribution), therefore logarithmic bins are used. All empirical distributions are displayed on fig.2 except for users with ID 14, 17 and 18 because they started less than 15 calls. We can conclude that different users of the same strength have similar interevent time distributions. Common properties of the interevent time distribution:

- interevent times less then 10 seconds are rare;
- P_{ie} starts with a small peak (with width$\sim 50-100$s);

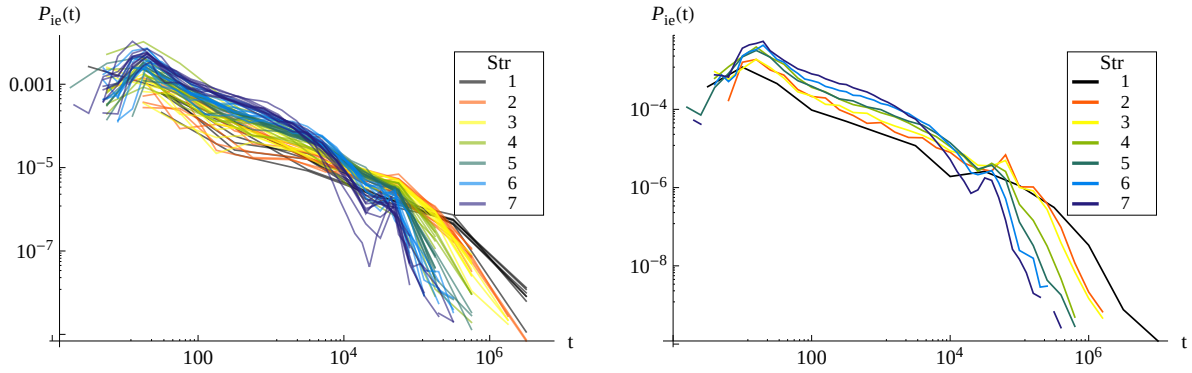


Figure 2: Empirical interevent time distribution for the whole sMPC dataset (individuals with ID $14, 17, 18$ have been excluded because they started less than 15 calls). The right panel shows the averaged interevent time distribution for the groups.

- there is a peak centered at about 10-24 hours (position change with the strength group, width of the peak is $\sim$ 3h).

The shape and the precise position of the peaks can be different for users in the same group (fig.3). Therefore averaging blurs the peaks, that is why the second peak is so small in Karsai's measurements [20]. In the case of users with low activity averaging helps to find the peak because smaller bins can be used. For the most active users there is a valley before the second peak. Fig.3 shows histograms with small bin sizes for active users and a gap in the interevent time is clearly visible for them. The cutoff before the gap is approximately exponential. For less active users the gap vanishes because they can produce large interevent times at daytime. The gap and the peak are

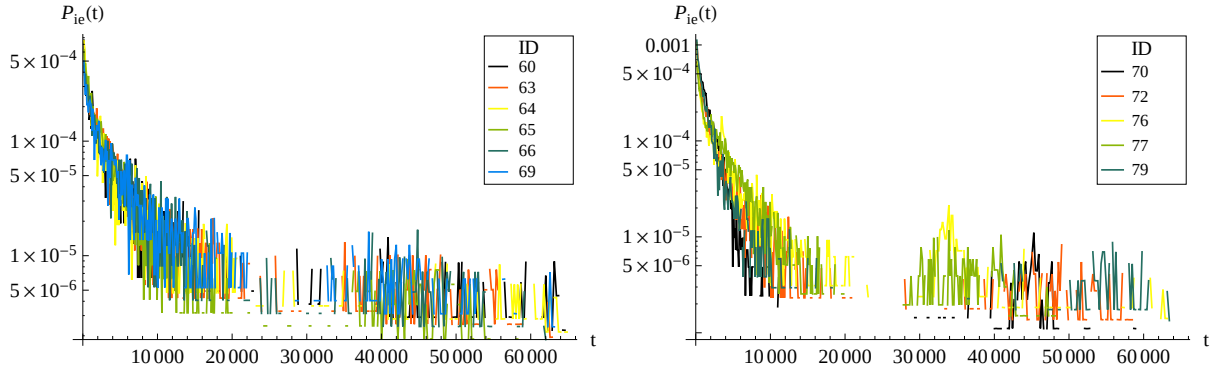


Figure 3: Interevent time histograms for active users with small bin sizes. The users on the plot have a gap in their interevent time distribution, but the position and the width of the gap depends on the user. The cutoff before the gap is approximately exponential.

also visible on the $\tau\tau$plots of the following chapter (fig.4,5). To analyse the peak and the valley one should use bins centered around the peak instead of simple logarithmic or uniform bins.

The section between the first peak and the valley can be different for the users in the same group. In most cases this section is nearly linear in a log-log plot with slope around -0.6 (those dominate the picture in fig.2). In a few cases a plateau with small or zero slope follows the first peak (ID 67, 68, 70). For these people the cascading nonhomogeneous Poisson process can make a good fit (the details can be found in section 3).

Correlations between interevent times

Interevent time distribution shows nothing about the present correlations. I calculated the correlation function of the interevent times:

$$C_{\tau\tau}(j) = \frac{\langle \tau_i \tau_{i+j} \rangle_i - \langle \tau_i \rangle_i^2}{\langle \tau_i^2 \rangle_i - \langle \tau_i \rangle_i^2} \approx \delta_{0,j}$$

This quantity doesn't show correlations between the interevent times. This is not a good measure for correlations in this case because of the broadness of the interevent time distribution.

Another way to see if there is some correlation between the consecutive interevent times is to plot the (τ_i, τ_{i+1}) pairs (where the actual interevent time is the abscissa, and the next is the ordinate). Let us call this $\tau\tau$plot. Correlations could be seen from the distribution of points in the plot. E.g. if the user liked to repeat the last interevent time, then the $\tau\tau$plot would be centered on the diagonal. If large τ were followed by small and vice versa, then the diagonal would be sparse. If the interevent times tended to be smaller than the last one, then the $\tau\tau$plot would be centered under the diagonal. We don't see anything like these in the empirical series, in fact, the $\tau\tau$plot does not change significantly if we randomize (permute) the interevent time sequence.

The $\tau\tau$plot clearly shows the peaks in the interevent time distribution (they appear as dense horizontal and vertical lines). For the first and second strength group it can be clearly seen that the interevent time distribution have peaks at integer times a day. As the user activity gets higher the large interevent times vanish (fig.4).

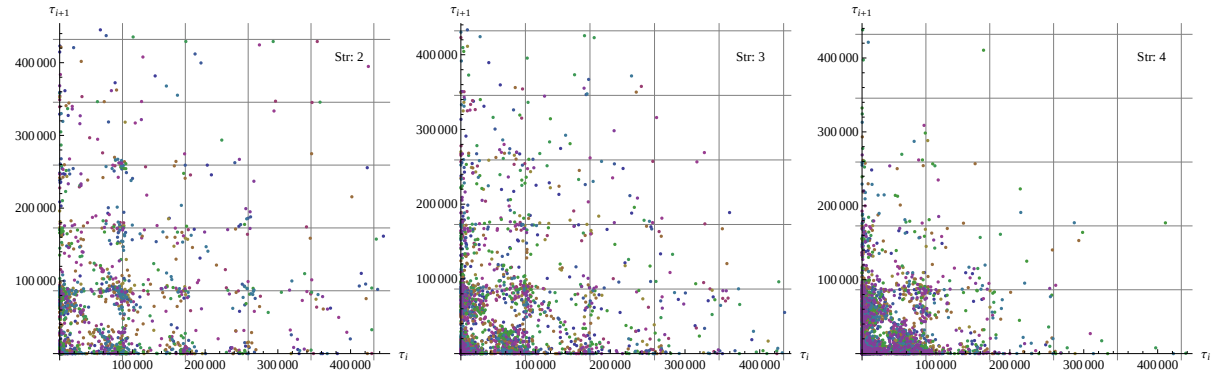


Figure 4: $\tau\tau$plot for stregth groups 2, 3 and 4. Coordinates of the points are consecutive interevent times. The grid shows the days.

In case of the most active users, there are very few interevent times greater than one day and the $\tau\tau$plot is similar for the users as well (fig.5).

2.3 Autocorrelation function

Another characteristic property is the autocorrelation function of the calling sequence. We define the indicator variable $X(t)$ which takes value 1 for the times an event occurs and 0 otherwise. The autocorrelation function is a 2-point function that depends on both variables because there is no time translation symmetry.

$$A(t,t') = \frac{\mathbb{E}[X(t+t')X(t')]}{\langle \mathbb{E}[X(s)] \rangle_s}, \tag{3}$$

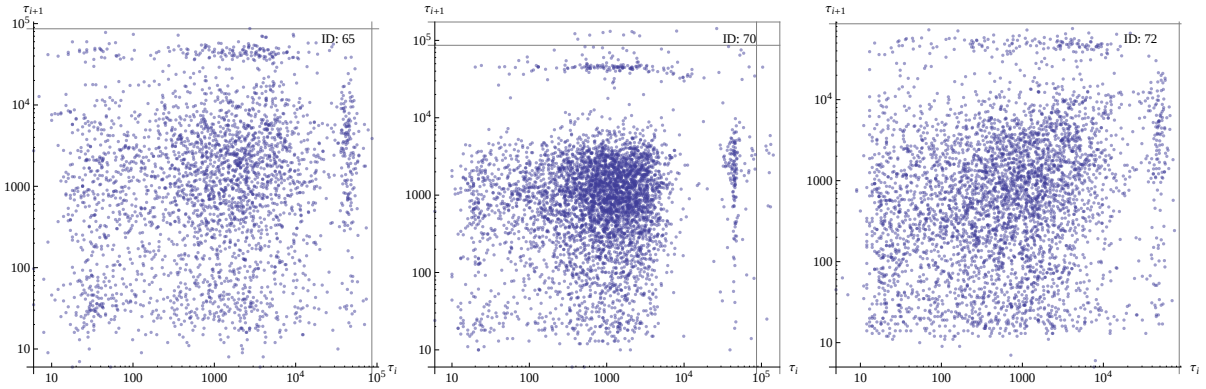


Figure 5: $\tau\tau$plot for users 65, 70 and 72. Coordinates of the points are consecutive interevent times, axes are logarithmic. The grid shows the days.

where the bracket symbol $\langle \dots \rangle_x$ means averaging in the variable in the lower index and the denominator is the average rate of calls. This function is hard to measure on individuals, especially on users with low activity. The quantity that is easy to measure is the conditional probability of the event: the user starts a new call t seconds after the last call:

$$A(t) = \mathbb{P}(\text{User starts a call } t \text{ seconds later} \mid \text{he starts a call now}) \tag{4}$$

We note that in the sMPC dataset the start and endpoint of the calls are the same (we don't care about the length of the calls). We use the name autocorrelation function for this conditional probability.

Proposition 1. *$A(t)$ is the average of $A(t,t')$ in the second variable.*

Proof. Let t_0 denote the random time of the actual call:

$$A(t) = \sum_s \mathbb{P}(X(t+s) = 1 | X(s) = 1)\mathbb{P}(t_0 = s) \tag{5}$$

and

$$\mathbb{P}(t_0 = s) = \frac{\mathbb{P}(X(s) = 1)}{\sum_s \mathbb{P}(X(s) = 1)} \tag{6}$$

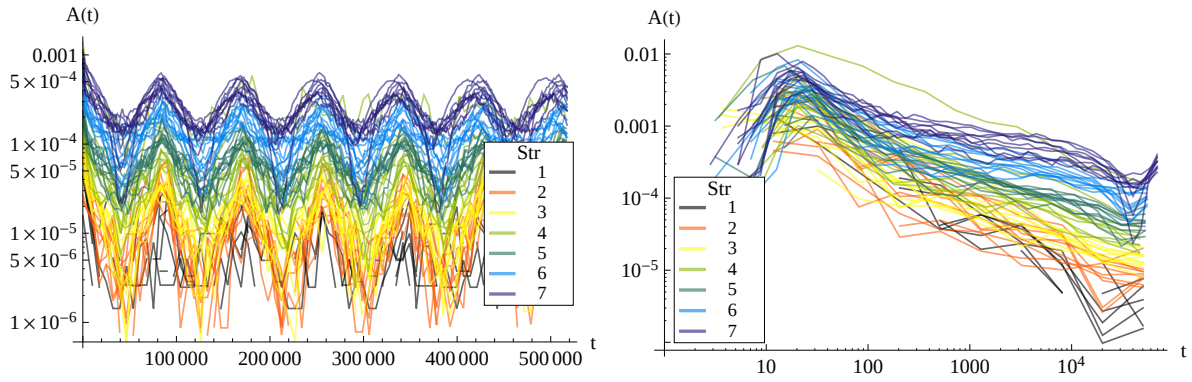
yields

$$A(t) = \frac{\sum_s \mathbb{P}(X(t+s) = 1, X(s) = 1)}{\sum_s \mathbb{P}(X(s) = 1)} \tag{7}$$

$$= \frac{\langle \mathbb{E}[X(t+s)X(s)] \rangle_s}{\langle \mathbb{E}[X(s)] \rangle_s} = \langle A(t,s) \rangle_s \tag{8}$$

For stationary processes $A(t,t') = A(t,0) = A(t)$. □

In this normalization $A(0) = 1$, and the asymptotic or average value of the autocorrelation function is not fixed to zero. In the empirical measurements the expectation value is left (compare (8) with (1)). The expectation value could be executed either if we assumed some periodicity in the calling sequence, but it would not lower the noise, or if the users were assumed to be taken as different samples from the same process.



(a) Autocorrelation on long range

(b) Autocorrelation on short range

Figure 6: Autocorrelation function for all users in the sMPC dataset. Panel (a) shows the long time and (b) the short time behaviour.

2.3.1 Autocorrelation function for long times and the circadian pattern

The empirical autocorrelation functions behave well in the sense that they are smooth for the users with high activity. For almost every user $A(t)$ is periodic with a period of a day (fig.6.(a)). To be more precise, the period differs from the length of a day due to the eliminated lengths of the calls. In the nonhomogeneous Poisson process the autocorrelation function can be calculated from the rate function (details are shown in section 3). The rate function is the probability density function of starting a call at a specific time of the day. This can be measured by projecting the call sequence into one day and estimating the density of the calls as a function of the projected timings. In the sMPC dataset it is hard to do the projection because the proper event sequence cannot be reproduced from the data due to the finite missing lengths of the calls. However we can still *visualize* the rate function or the *circadian pattern* with the following method.

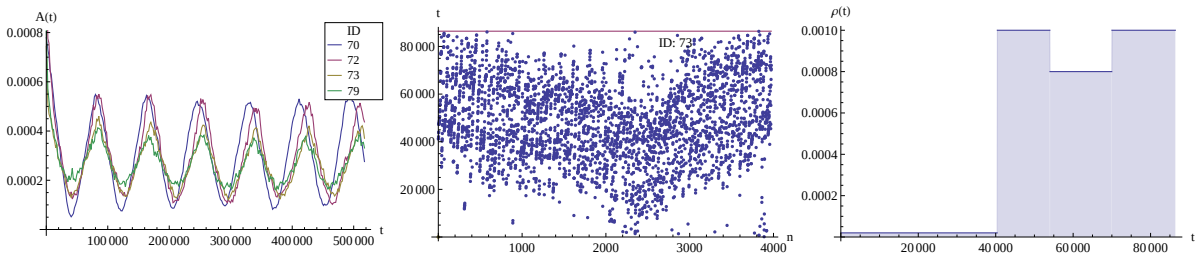
Circadian plot We determine the event sequence neglecting the length of the calls. Now the length of the days is smaller than 24 hours and is not equal for different days either. We assume that there exists an effective length of the days for the users. Now we do the projection: we take the modulo of the event sequence with this length. In the end we plot the projected timing of events as a function of the consecutive events.If we choose a proper effective length and length of days do not fluctuate much, the circadian pattern appears as denser and sparser vertical bands (fig.7-9.b). The rate function would be the density along the time axis (this is an averaged rate function over the whole length of the data). Due to the fluctuations in the missing length of the days the circadian plot is not translation invariant (fig.7.b) but still shows the circadian pattern. The rate function could be measured if we detrended the data on the plot. I have not done it because the rate function could be determined easier from the MPC dataset. If the user changed his circadian pattern during the examined period, the circadian plot would show it. The sMPC dataset shows that the circadian pattern is robust (we note here that in contrast to this the average daily activity can change on the scale of 100 days).

Remark. *The method used to construct the circadian plot allows one to estimate the mean and variance of the duration of calls from a dataset which does not contain information about the length of calls. The necessary condition to do so is the existence of the circadian pattern.*

I noticed that similar autocorrelation functions are connected to similar rate functions. I identified four main classes of the autocorrelation functions, though there are many users who cannot be classified into these classes.

Sawtooth function (fig.7)

The autocorrelation function of users with ID 60, 70, 72, 73, 79 is similar to the sawtooth function. The rate function is nearly homogeneous for these people, and the length of the active period is about half a day. Users with ID 77, 78 can also be appended to this list but they have an extra small peak where others have the minima (this comes from their longer active period).

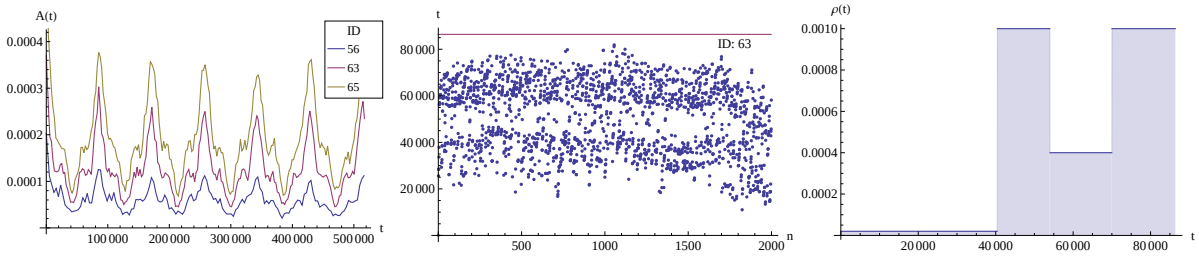


(a) Autocorrelation function (b) Circadian plot (c) Rate function

Figure 7: (a) Sawtooth autocorrelation function of some users. (b) Event sequence projected to one day, the number of the call is on axis n and the projected timing is on axis t. (c) Typical rate function in the group.

Peak with shoulder (fig.8)

This is the most populous group. The autocorrelation function has small peaks on both sides of the daily peaks. The following users are in this group: ID 23, 27, 28, 29, 37, 38, 40, 42, 44, 46, 48, 51, 52, 56, 63, 65. All of them have two peaks in their rate function, similar to one in fig.8.(c). Users with ID 66 and 67 have also similar autocorrelation function, but their rate function is different (also from each other).

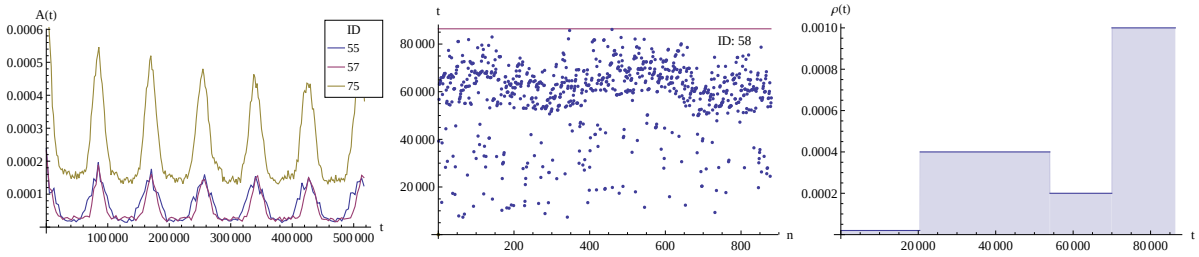


(a) Autocorrelation function (b) Circadian plot (c) Rate function

Figure 8: (a) Autocorrelation functions with peaks on the shoulders. (b) Event sequence projected to one day, the number of the call is on axis n and the projected timing is on axis t. (c) Typical rate function in the group.

Plateau between the peaks (fig.9)

The autocorrelation function has a nearly constant section between the peaks. The following users are in this group: ID 21, 22, 55, 57, 75. For the users with higher activity (ID 55,57,75), the typical rate function is shown in fig.9.(c). The essential property for the group is the narrow and very active section in the rate function. Users 21 and 22 have only this narrow peak in their rate function.

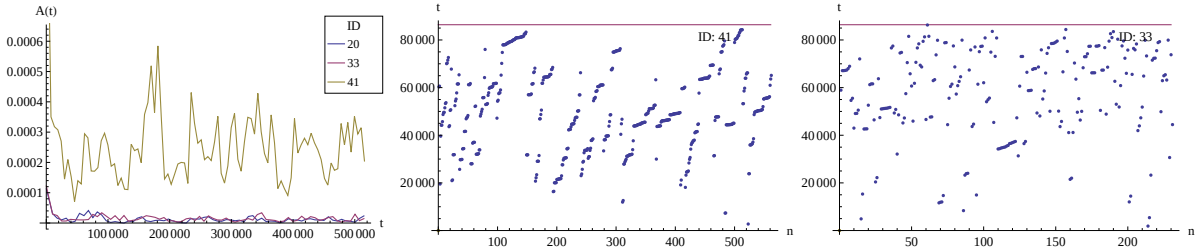


(a) Autocorrelation function (b) Circadian plot (c) Rate function

Figure 9: (a) Autocorrelation functions with plateau between the peaks. (b) Event sequence projected to one day, the number of the call is on axis n and the projected timing is on axis t. (c) Typical rate function in the group.

Nonregular or bursty behaviour (fig.10)

There are some users whose autocorrelation function is not regular (ID 20,33,41). Their phoning activity is directed by bursty periods that appear as lines with small slope in the circadian plot in fig.10.(b-c).



(a) Autocorrelation function (b) Circadian plot (c) Circadian plot

Figure 10: (a) Nonregular autocorrelation functions. (b-c) Event sequence projected to one day, the number of the call is on axis n and the projected timing is on axis t.

Summary

The long time region of the autocorrelation function is determined by the rate function. This is verified by simulations and some examples are shown when the cascading nonhomogeneous Poisson process is studied (section 3). The rate function of most users can be modelled by four constant sections. The long section with very low activity probably corresponds to the nighttime.

There are two daily peaks and a valley between them. The shape of the autocorrelation function depends on the concrete parameters of these four domains (their width and height), therefore the classification discussed is not sharp.

2.3.2 Short time autocorrelation and burstiness

The empirical short range autocorrelation functions for the whole sMPC dataset are shown in fig.6.(b). The most conspicuous property of autocorrelation functions is that they start with a peak. This peak cannot emerge due to the circadian pattern because the rate function is assumed to be smooth.

Our aim is to detect and analyse the bursty behaviour. To do this, we have to find a reference where the burstiness is not present but the circadian pattern is (the average value tells nothing because there is a structure above that). Fortunately the autocorrelation function is smooth enough, so we can compare the short time autocorrelation function with the *autocorrelation function shifted by one-two-... days* at the same range. The peak characteristic for the short time behaviour is expected to be absent in the shifted autocorrelation functions and it is verified by the measurements.

The reference curve is the average of the autocorrelation functions shifted by $1,2,\dots 9$ days of length T_{eff}:

$$A^{ref}(t) = \frac{1}{9}\sum_{i=1}^{9} A(t + i\,T_{eff})\,. \tag{9}$$

The only problem with this reference curve is that it is not normalized properly. The weekends (the periodically appearing days with lower activity) have small contribution to the first day of the autocorrelation function because the number of calls is small on these days. But the weekends lower the autocorrelation function for times larger than a day. This can be handled, but it should be done in the large SMC dataset where the noise due to the effective length of the days is not present. The end of the peak is where the slope of the short time autocorrelation reaches the slope of the reference line. In some cases the peak is narrow (its width is $\lesssim$ 1000s, eg. ID 62, 66, 67, 68, 70, 73, 75, 77), while in other cases the peak is thick (its width is $\gtrsim$ 10000s, eg. ID 61, 64, 65, 72, 78, 79). Some examples of the short range autocorrelation function are shown in fig.11.

2.4 Distribution of the number of bursty events

As it was mentioned in the introduction, a new measure for dependencies within the bursty period was recently introduced. This is the *distribution $P_{E,\Delta t}(n)$ of the number of events E in a bursty period defined for time window Δt*. Two subsequent events are in the same bursty train (period) if the time elapsed between them is smaller than the time window. This quantity decays exponentially in variable n for any Δt for renewal processes (eq.(2)). Márton Karsai's measurements show that for the MPC dataset, this quantity has a power-law regime and the exponent ($\gamma \approx 4.1$) does not depend on the time window when it is chosen from the interval between 32 seconds and 9 hours. This measurement shows an average and it should be clarified that this regime exists for all individuals or it comes from some individuals or a group of them. For this reason Márton Karsai also measured the P_E distribution for the different strength groups, and he found power-law regimes with exponents decreasing in activity ($\gamma = 4.3$ for the least active and $\gamma = 2.45$ for the most active group) [20].

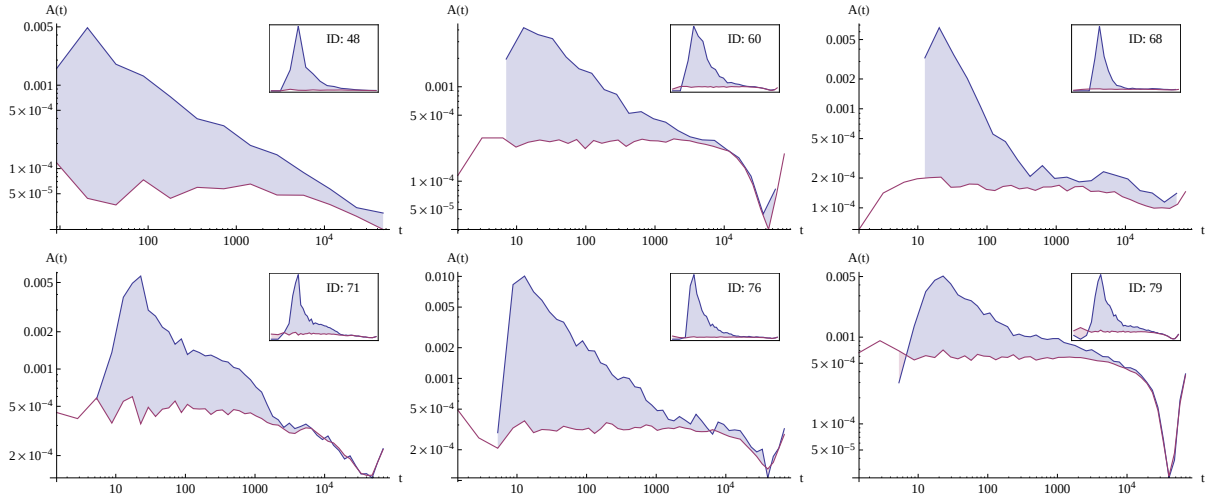


Figure 11: Short range autocorrelation for some users (axes are logarithmic). The violet (lower) curve is the reference. In the inset only the horizontal axis is logarithmic and the vertical is linear.

Measurements in the sMPC dataset

The 180 days of the extended MPC and the sMPC datasets seem to be too short to get a clear picture on the power-law regime of the P_E distribution, but it is also important to see every quantity on the individual level.

The first question is the dependence on the time window. For $\Delta t < t_{c,l} \sim 8$s every event is separated and defines a bursty train, therefore $P_E(n) \approx \delta_{n,1}$. The other limit is $\Delta t > t_{c,u} \sim 100$days when the P_E distribution is centralised on the overall activity of the user. Between these two limits P_E shows nontrivial behaviour. When $\Delta t\,(> t_{c,l})$ is small, the P_E distribution is monotonically decreasing and the decay is fast.

Increasing Δt results in appearing larger values of E. More and more bursty trains merge and the days are cut only to few trains (fig.12). When $\Delta t \approx 9$h, the trains cover the days for almost all users and the P_E distribution gets similar to the distribution of the numbers of daily calls. They are not the same because a call at night can merge two days into one train or a day can be split into two trains by a long inactive period at daytime, and furthermore the P_E distribution does not count the days with no calls. If we increase Δt further, days start to merge and only days with low or no activity can split the trains.

If Δt is small, P_E is capable to measure short-range dependencies deeper than the autocorrelation function. If Δt is large ($\gtrsim 1$day), then P_E gives some information on the occurrence of periods with low activity.

Fig.13 shows that the beginning section of the P_E distribution can be well approximated by an exponential function (i.e. $\log P_E$ is approximately linear in n) when $\Delta t < 1h$. When $\Delta t \lesssim 512s$, the large trains are very rare but we should not find any if the P_E was exponential. The queuing models try to take this into account as an intrinsic property of the model, while in the nonhomogeneous Poisson process these long trains may come from some external effect (e.g. birthday, ...) or from a large scale change in the phoning activity. These rare events can produce the power-law regime in Márton Karsai's measurements after averaging between the strength groups. We have to note that the strength groups cover wide range in the activity and the effect of blank periods, which we will introduce in the following subsection, can pull active users to a less active group. This is

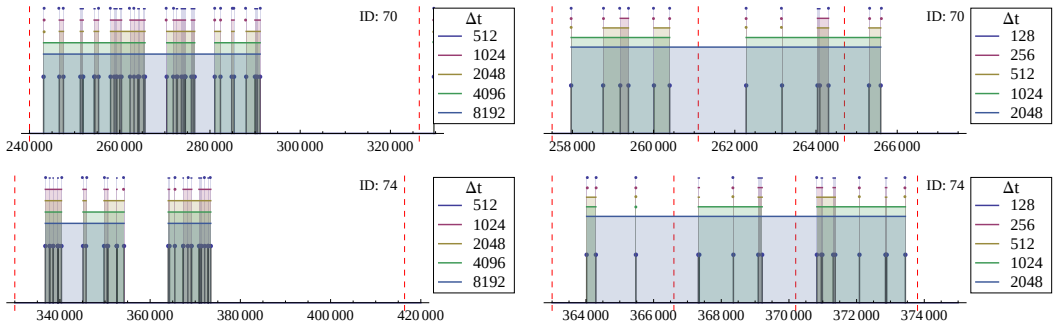


Figure 12: Event sequence and bursty trains for some Δt values. The grid shows days on the left and hours on the right images.

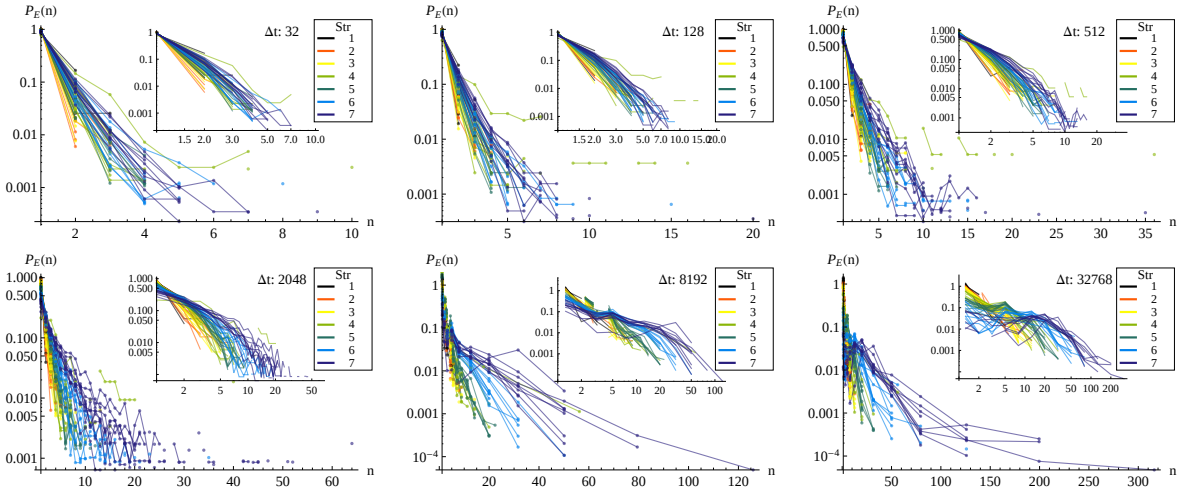


Figure 13: P_E distribution for the whole sMPC dataset for some time windows. The main plots are semi-logarithmic, and the insets show the logarithmic plot of the same data. Logarithmic binning is used in the case of the two largest time windows.

another possible candidate for the rare events in the tail of the P_E distribution.

2.5 Calls in daily resolution

The method used to visualise the circadian pattern can also be used to determine the boundary of the days from the interevent time sequence. After doing this, we can analyse the number of calls started on different days $D(n)$, where n counts the days. The following quantities could be analysed:

- distribution $P_D(n)$ of the number of calls (n) started on a day
- the homogeneity on large scales
- periodicity on weekly or monthly scales

Empirical distribution function

The empirical distribution functions are shown in fig.14. Nonhomogeneous Poisson processes

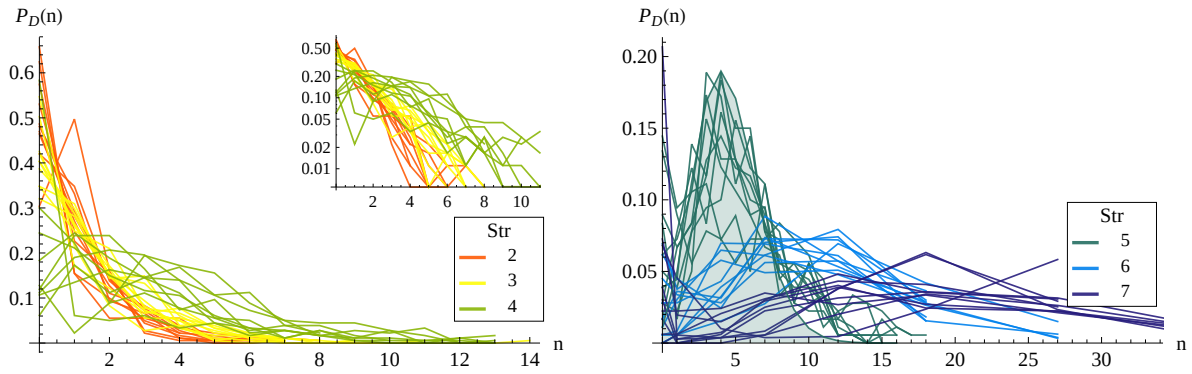


Figure 14: Empirical probability density function of number of calls started on a day. The filled area on the right plot shows a Poisson distribution with parameter $\lambda = 4.5$.

would result in Poisson distributed number of calls for a day if the rate function was the same for every day. If there were periodically occurring days with lower or higher activity, we would find a mixture distribution of Poisson distributions with different parameters. This makes the fitting of Poisson distribution to the data more complicated.

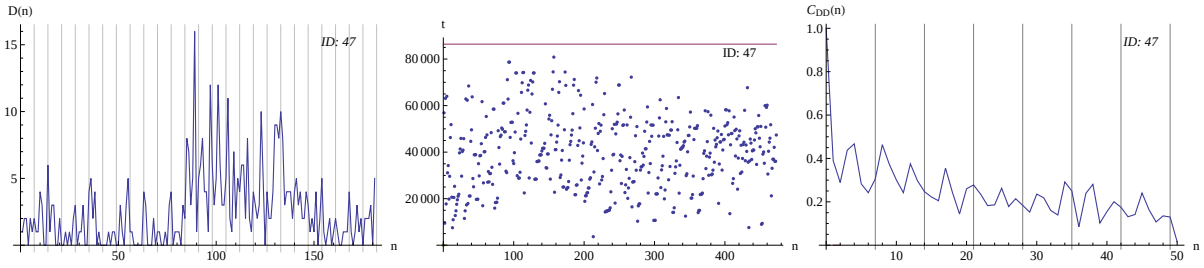
Inhomogeneity on large scales

We usually think that users behave in a stationary manner. It is an interesting question how good this assumption is. Analysing the number of daily calls may help to decide this. More than half of the users seem to behave rather stationary. Their daily call function fluctuates around its mean and there are some spikes. In contrast with this there are many users whose daily call function seems to change in time. These users can be divided into two (not essentially disjoint) groups. The first group contains the ones (ID: 11, 17, 21, 22, 27, 29, 39, 41, 43, 79) who had a very long inactive period (e.g more than 25 days). This criterion is fulfilled in the entire strength group 1, but many of the members of this group have these gaps homogeneously, so they should not be added to this group. The second group contains the users who seem to have modified their activity along the 180 days but not to zero (ID: 35, 42, 43, 47, 55, 60, 62, 73, 76, 78). These changes can also be seen on the circadian plot because the change in the activity implies change in the effective length of the days (fig.15.(a-b)).

Remark. *User 41 used his telephone only in the first 25 days, but has made 561 calls. With this performance he/she should have been assigned to the most active group (instead of group four based on the average activity). These types of blank periods should be taken into account in an accurate modelling.*

Periodicity on weekly scale

As shown before, users have a well-defined circadian pattern. Can we see a similar weekly pattern? The answer is ambiguous. One method to visualise a large scale (x-day) periodicity is to calculate



(a) #of daily calls (b) Circadian plot (c) Correlation between daily calls

Figure 15: (a) The user switches to higher activity between days 85 and 150. The grid shows the weeks (also for panel (c)). (b) The first region with positive slope corresponds to longer effective length of the days due to the lower activity. (c) The first 3 peaks suggest a 4-day periodic behaviour.

the autocorrelation function of the daily calls.

$$C_{DD}(n) = \frac{\langle D(m)D(n+m)\rangle_m - \langle D(m)\rangle_m^2}{\langle D(m)^2\rangle_m - \langle D(m)\rangle_m^2} \tag{10}$$

I found that most users do not show periodic large scale patterns. For them every day seems to be equal, i.e the days with lower or higher activity than average occur in a nonperiodic manner. 15 users (ID: 28, 38, 40, 45, 48, 56, 58, 60, 62, 63, 68, 69, 74, 76) show nice, quasi-periodic autocorrelation functions with peaks at $7-14-21-\ldots$ days. There are some other users (ID: 24, 26, 44, 52, 55, 64, 66, 75, 77) who show these peaks but the correlation is smaller and noisier for them. There are 3 users (ID: 39, 47, 50) who have peaks at integer times 4 days (fig.15.(c)). User 70 is an interesting exception because he/she has a period of two weeks (1 day inactive, 7 days active, 1 day inactive, 5 days active). Fig.16 shows some examples for nice periodic daily call correlation functions, and fig.17 shows some counterexamples.

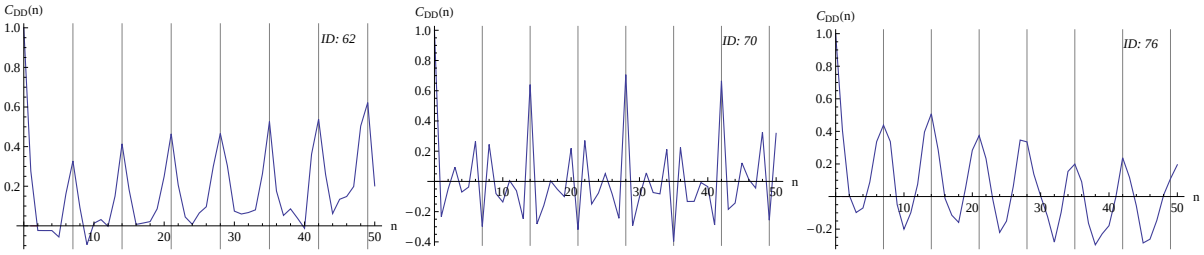


Figure 16: Daily call correlation function. Grid shows the weeks.

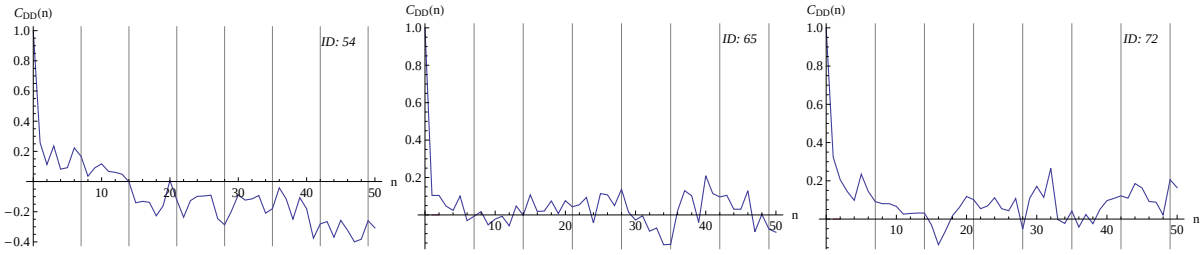


Figure 17: Daily call correlation function. Grid shows the weeks.

3 Nonhomogeneous Poisson process

In [12] people are assumed to send emails according to a cascading nonhomogeneous Poisson process. In their model a lot of parameters have to be fitted because the model needs both the rate function and the distribution of the number of events in a cascade. They did not analyse their model in some simple cases that would help to understand the model and its consequences.

For the sake of simplicity first we study the nonhomogeneous Poisson process without cascades, and we calculate some quantities that were analysed on the datasets. In the second half of this section I will show some simulations for the characteristic functions of the cascading nonhomogeneous Poisson process.

Let us denote the rate function by ρ. The probability of starting a call in an interval $(t,t+dt)$ is equal to $\rho(t)dt$. The rate function is assumed to be periodic with period of T, i.e. $\rho(t+T)=\rho(t)$, where T can be a day or a week, etc.

3.1 Interevent time distribution

Let $P_{>,x}(t)$ denote the probability that no event happens from time x up to $t+x$.

$$P_{>,x}(t) = e^{-\int_x^{x+t}\rho(t')\mathrm{d}t'} \tag{11}$$

The probability distribution function of interevent times can be calculated by conditioning on the timing of the actual call (similarly to the train of thought used in computing the autocorrelation function in eq.(5)-(6)):

$$\mathbb{P}(\tau<t) = 1-\mathbb{P}(\tau>t) = 1-\frac{\int_0^T \rho(x)e^{-\int_x^{x+t}\rho(t')\mathrm{d}t'}\mathrm{d}x}{\int_0^T \rho(t')\mathrm{d}t'} \tag{12}$$

The probability density function $P_{ie}(t)$ of the interevent times:

$$P_{ie}(t) = \frac{\int_0^T \rho(x)\rho(x+t)e^{-\int_x^{x+t}\rho(t')\mathrm{d}t'}\mathrm{d}x}{\int_0^T \rho(t')\mathrm{d}t'} \tag{13}$$

From this formula it is easy to see that the interevent time distribution inherits the periodicity of the rate function:

$$P_{ie}(t+T) = e^{-\int_0^T \rho(t')\mathrm{d}t'}P_{ie}(t) \tag{14}$$

which means that the integral in (13) have to be calculated only for $t\in[0,T]$.

General squarewave rate function

To get more insight into the nonhomogeneous Poisson process, we analyse the case when the rate function is a squarewave function. Let N denote the length of nighttime and D the length of daytime ($T=N+D$).

$$\rho(t) = \begin{cases} a & \text{if } t\in(0,N), \\ b & \text{if } t\in(N,N+D), \end{cases} \tag{15}$$

and $\rho(t+T)=\rho(t)$. Without loss of generality we assume that $N<D$. Evaluating (13) for $t<N+D$,

$$P_{ie}(t)=\begin{cases} a^2e^{-at}(N-t)+b^2e^{-bt}(D-t)+ & \\ +\frac{2ab}{b-a}\left(e^{-at}-e^{-bt}\right) & \text{if } t\in(0,N), \\ b^2e^{-bt}(D-t)+b^2e^{-bt}e^{(b-a)N}(t-N)+ & \\ +\frac{2ab}{b-a}\left(e^{(b-a)N}-1\right)e^{-bt} & \text{if } t\in(N,D), \\ a^2e^{(b-a)D}e^{-at}(t-D)+b^2e^{(b-a)N}e^{-bt}(t-N)+ & \\ +\frac{2ab}{b-a}\left(e^{(b-a)N}e^{-bt}-e^{-(b-a)D}e^{-at}\right) & \text{if } t\in(D,N+D), \end{cases} \tag{16}$$

For $t>T=N+D$ we can use the formula below

$$f(t+T)=e^{-(aN+bD)}f(t) \tag{17}$$

3.2 Autocorrelation function

The autocorrelation function is the conditional probability of starting a call t seconds later if a call is started now.

$$A(t)=\frac{\int_0^T\rho(t')\rho(t+t')\mathrm{d}t'}{\int_0^T\rho(t)\mathrm{d}t} \tag{18}$$

It is easy to see that $A(t+T)=A(t)$, i.e. the autocorrelation function is periodic. This can be evaluated on some examples:

General squarewave rate function

Let N denote the length of nighttime and D the length of daytime.

$$\rho(t)=\begin{cases} a & \text{if } t\in(0,N) \\ b & \text{if } t\in(N,N+D) \end{cases}$$

Without loss of generality we assume that $N<D$. It is easy to evaluate (18) for $t<N+D$,

$$A(t)=\frac{1}{aN+bD}\begin{cases} a^2(N-t)+b^2(D-t)+2abt & \text{if } t\in(0,N) \\ 2abN+b^2(D-N) & \text{if } t\in(N,D) \\ a^2(t-D)+b^2(t-N)+2ab(D+N-t) & \text{if } t\in(D,N+D) \end{cases} \tag{19}$$

For $t>T=N+D$ we can use the periodic extension. This expression for the autocorrelation function can explain the emergence of the plateaus in fig.9.a . That is if the rate function is dominated by a peak and on the other regions it is negligible, then $A(t)$ is constant when $t\in(D,N)+kT$, $k\in\mathbb{N}$ (in this case $D<N$).

General periodic function

The rate function ρ is assumed to be periodic (with period T), so it can be written as Fourier series:

$$\rho(t) = \sum_{n=-\infty}^{\infty} a_n e^{i\omega nt} \tag{20}$$

where $\omega = \frac{2\pi}{T}$ and

$$a_n = \frac{1}{T}\int_0^T \rho(t) e^{-i\omega nt} \mathrm{d}t \tag{21}$$

Substituting the Fourier series representation of ρ into (18) yields

$$A(t) = a_0 + \frac{1}{a_0}\sum_{n=1}^{\infty} |a_n|^2 \, 2\cos(n\omega t) \tag{22}$$

This formula clearly shows how the rate function determines the autocorrelation function. It also says that the autocorrelation function is not enough to calculate the rate function, because it does not contain information about the phases of coefficients in the Fourier series.

Number of bursty events

The distribution of the number of bursty events is very hard to calculate in the nonhomogeneous Poisson process. It could be investigated by simulations but the aim of the first half of this section is to see some simple analytic formulae for the characteristic functions.

3.3 Cascading nonhomogeneous Poisson process

In the cascading nonhomogeneous Poisson process every event of the main process triggers a second Poisson process with rate λ. The length of the second process is given by the number of events N_E which is chosen from some distribution $P_N(n)$. If we consider the cascade process as a small perturbation, then the interevent time distribution is a mixture distribution of the interevent time distribution of the cascade process and of the nonhomogeneous Poisson process. The weights are given by expectation value of N_E.

$$P_{ie}^{CNHP}(t) = \frac{P_{ie}^{NHP}(t) + \mathbb{E}[N_E] P_{ie}^{C}(t)}{1 + \mathbb{E}[N_E]} \tag{23}$$

where the upper indices refer to the process (C: cascade, NHP: nonhomogeous Poisson). This expression fits the simulations very well (fig.18 and 19). This formula shows that the cascade process brings in one more scale and a simple Poisson process as a cascade cannot reproduce the interevent time distributions for many users seen in the sMPC dataset.

The cascade process also gives a peak in the autocorrelation function. The contribution of the cascade process to the autocorrelation function can be given by the following formula:

$$A^C(t) = \sum_{n=1}^{\infty} \mathbb{P}\left(\sum_{i=1}^{n} \tau_i^C = t\right) \mathbb{P}(N_E > n) \tag{24}$$

where τ_i^C are the interevent times of the cascade process. This expresses that the call at time t can be the first, second, ... The autocorrelation function for short times can be approximated by

$$A^{CNHP}(t) \approx A^C(t) + A^{NHP}(t) \tag{25}$$

Comparison of the model and measurements

By changing the values of the model parameters I have studied how good fits the model can produce. This short section is just a demonstration because the fits are not optimal. Making fits by mathematical methods is planned to be done in the future. The cascade process I have used here is a modified version of the Poisson process (which is no more a Poisson process but similar to it): every interevent time is increased by a constant value x to take the reaction time into account. The simulation results are shown in fig.18 and fig.19.

The cascade Poisson process is capable of producing a peak with an exponential decay for both

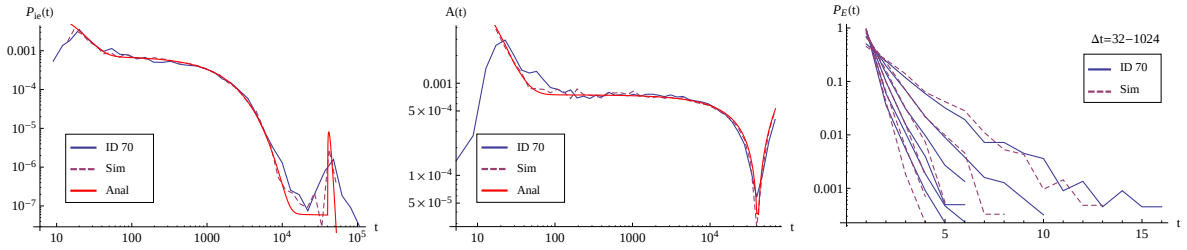


Figure 18: Characteristic functions for user 70 and for the cascading Poisson process with parameters fitted by hand. The analytic approximating curves were calculated from (23) and (25).

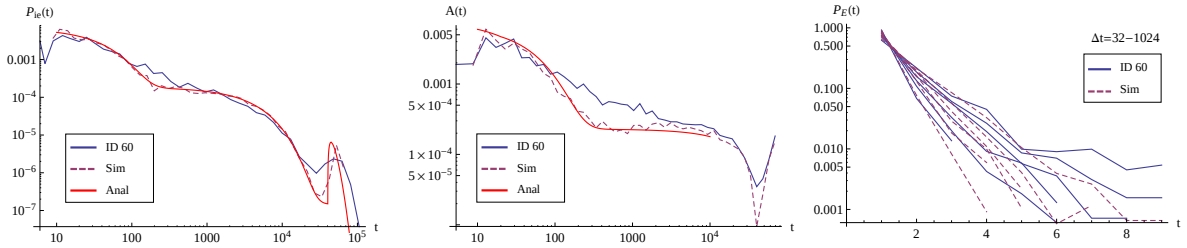


Figure 19: Characteristic functions for user 60 and for the cascading Poisson process with parameters fitted by hand. The analytic approximating curves were calculated from (23) and (25).

of the interevent time distribution and the autocorrelation function (fig.18), but it cannot reproduce an approximately scale-free decay which can be seen in fig.19 (and in fig.11 for users 48, 60, 78, 79).

As a summary we conclude that the nonhomogeneous Poisson process can make a very good fit for the autocorrelation function for long times, but this can be reproduced by any process with a rate function (this statement is based on simulations which we do not show here). The Poisson cascade can make a good fit in the short time region for some users who have an exponentially decaying peak in their interevent time distribution but for many others it does not work well.

There are two ways to continue the modelling. The first way is to study processes with fat tailed interevent time distribution as cascades which are triggered by a nonhomogeneous Poisson

process. The second way is to analyse the effect of introducing a rate function to a process with slowly decaying autocorrelation function and interevent time distribution.

In the following we will study simple processes leading to slowly decaying autocorrelation functions.

4 Priority arranged list model

Task queuing models are studied in human dynamics because they are capable of producing power-law regime in the interevent time distribution. We present a model which can produce a power-law decay also in the autocorrelation function.

We assume that every person has a task list and people choose their forthcoming activities from the list. The list is ordered in the sense that the probability of choosing the i^{th} activity from the list is monotonically decreasing as a function of position i. The chosen activity is going to be executed and *it jumps to the front of the list*. This means that if a person uses his phone at one time, phoning as an activity is going to have high priority for a while. This mechanism is responsible for producing the bursty behaviour.

The model is capable of covering many types of activities at one time. We classify the activities into two groups. The important activities – the phoning tasks – will be denoted by A and all other activities will be denoted by B (the latter form the group of "unimportant" activities). The length of the list is denoted by N and the quantity of activities type B is denoted by L, i.e. we have $N-L$ activities of type A. The following example shows a realisation of a list and two transitions (the chosen activities are coloured in each step):

$$ABAABBAA\mathbf{A}AB...$$
$$AABAA\mathbf{B}BAAAB...$$
$$BAABAABAAAB...$$

In the first step the user started a call and after that he started some other activity.

We note that the time evolution of the first element of the list gives the event sequence of the user. We are going to study the interevent time distribution, the autocorrelation function and the distribution of the number of bursty events for this model.

Relation to the theory of shuffling cards

We can define the time reversed version of the model in which we choose a position from the same distribution as in the original model and we move the first element of the list to the random position. This model has similar properties to the original model, e.g. this model has the same interevent time distribution and autocorrelation function. If we think of the list as a *deck of cards*, then the time reversed model is a generalisation of the *top-in shuffle* method. The latter is a primitive model of shuffling cards: the top card is removed and inserted into the deck at a uniform distributed random position.
We are going to study the original version of the model.

4.1 Stationary solution

The list model defines a Markov-chain. The state space is the space of permutations with repetition of sequences containing $N-L$ elements of type A and L elements of type B. Therefore the state space consists of $\binom{N}{L}$ elements. The probability of the next state depends on the current state only (this is the Markov property). The transition probability is determined by the probabilities w_i of choosing the i^{th} element. Here our only constraint is that for every position of the list $w_i \neq 0$ (we

are going to assume a power-law decaying function later). We arrange the probabilities of being in particular states to row vectors. The time evolution of this vector is generated by multiplying with the transition matrix from the right.

Example. $N = 4, L = 3$ *model*

The state space consists of 4 elements, and we are using the following ordering: $1 : ABBB$, $2 : BABB$, $3 : BBAB$, $4 : BBBA$. The transition matrix for this model:

$$W = \begin{pmatrix} w_1 & w_2+w_3+w_4 & 0 & 0 \\ w_2 & w_1 & w_3+w_4 & 0 \\ w_3 & 0 & w_1+w_2 & w_4 \\ w_4 & 0 & 0 & w_1+w_2+w_3 \end{pmatrix} \tag{26}$$

Proposition 2. *The transition matrix is doubly stochastic (sum of the rows and the columns are equal to one).*

Proof. We use an extended Markov chain for the proof in which the state space is the permutation of numbers $12\ldots N$ ($N!$ elements are in the state space). These numbers denote the initial positions of the elements of the list. $N-L$ numbers refer to type A and L numbers refer to type B. The Markov chain is ergodic, because all permutations can be generated by finite steps of moving one element to the front of the list (this is also true for the permutations of $AAA\ldots BB$). The sum of the rows are equal to one by definition (after each step we are surely in one of the states).
An arbitrary $p_1p_2p_3\ldots p_N$ permutation can be generated from itself with probability w_1, from p_2p_1 $p_3 ... p_N$ with probability w_2, ..., and from $p_2p_3\ldots p_Np_1$ with probability w_N, and from other permutations it cannot. The transition to a specific state is described by the columns of the transition matrix, therefore the sum of each column is $\sum w_i = 1$.

The row vector $(111\ldots1)$ is an eigenvector of a doubly stochastic matrix with eigenvalue 1, and this has to be the stationary solution of the chain because of ergodicity. Thus the stationary distribution is uniform, which means that after a long time every permutation is equally probable. From this it follows that the stationary solution of the original Markov chain (where the states are the permutations with repetition of $AA\ldots B$) is also uniform. This implies that the transition matrix is doubly stochastic. □

4.2 Master equation

The space of permutations or permutations with repetition consist of $N!$ or $\binom{N}{L}$ elements. If we are interested only in the $P_{A,i}(t)$ probability of finding an A at the i^{th} position of the list at time t then it can be described by a (row) vector of length N ($\underline{P}_A(t)$). The equation which describes the time evolution of this vector is called the Master equation. This equation can be determined if we use the law of total probability for $\underline{P}_A(t+1)$ with condition on the last step:

$$P_{A,1}(t+1) = \sum_{k=1}^{N} P_{A,k}(t) w_k \tag{27}$$

$$P_{A,k}(t+1) = P_{A,k}(t) \sum_{i=1}^{k-1} w_i + P_{A,k-1}(t) \sum_{i=k}^{N} w_i \tag{28}$$

The first line expresses that a task A from any position i can jump to the first place with probability w_i. The first term in the second line expresses that we can find an A at the $k^{th}(\geq 2)$ position if it has also been there in the step before and we have chosen a position smaller than k. The second term shows that it can also happen if an A was at the $k-1^{th}$ position, and we have chosen a task from position greater than k.

The Master equation can be written in a vectorial form:

$$\underline{P}_A(t+1) = \underline{P}_A(t)A \tag{29}$$

where the transition matrix

$$A = \begin{pmatrix} w_1 & \sum_{i=2}^{N} w_i & 0 & \cdots & 0 \\ w_2 & w_1 & \sum_{i=3}^{N} w_i & 0 & \vdots \\ w_3 & 0 & w_1+w_2 & \sum_{i=4}^{N} w_i & 0 \\ \vdots & \vdots & 0 & \ddots & w_N \\ w_N & 0 & \cdots & 0 & \sum_{i=1}^{N-1} w_i \end{pmatrix} \tag{30}$$

Remark. *In the $L = N-1$ model the transition matrix W of the Markov chain and A of the Master equation are equivalent if the ordering of the states are identical.*

Remark. *The Master equation can be derived from the transition matrix of the Markov chain.*

For many questions it is enough to study the Master equation with a transition matrix of size $N \times N$ instead of the transition matrix of the original Markov chain of size $\binom{N}{L} \times \binom{N}{L}$.

4.3 Stationary autocorrelation function

We are going to use the autocorrelation function with different normalisation for this model: we subtract the stationary value of the autocorrelation function and we normalise it. In terms of the indicator function of events $X(t)$ it reads

$$\mathscr{A}(t) = \frac{\mathbb{E}\left[X(0)X(t)\right] - \left\langle \mathbb{E}\left[X(t)\right]\right\rangle_t^2}{\left\langle \mathbb{E}\left[X(t)\right]\right\rangle_t - \left\langle \mathbb{E}\left[X(t)\right]\right\rangle_t^2}\,. \tag{31}$$

We can simplify this formula using proposition 1 for stationary processes:

$$\mathscr{A}(t) = \frac{\mathbb{P}\left(X(t)=1|X(0)=1\right) - \left\langle \mathbb{E}\left[X(t)\right]\right\rangle_t}{1 - \left\langle \mathbb{E}\left[X(t)\right]\right\rangle_t}\,. \tag{32}$$

This definition is more convenient to study stationary processes than (8).

The sequence of the executed tasks is given by the time evolution of the first element of the list. The probability of finding a task of type A in the first position of the list at time t is given by $P_{A,1}(t)$. The stationary distribution is uniform:

$$P_A(1, t\to\infty) = \frac{N-L}{N} \tag{33}$$

We use the following initial condition:

$$\underline{P}_A^{(L)}(t_0) = \left(1, \frac{N-L-1}{N-1}, \ldots, \frac{N-L-1}{N-1}\right) \tag{34}$$

because we are interested in the question how the list relaxes to the stationary state after a small perturbation.

$$\mathscr{A}^{(L)}(t) = \frac{P_A(1,t) - \frac{N-L}{N}}{1 - \frac{N-L}{N}} \tag{35}$$

Here we used the property that $\langle \mathbb{E}[X(t)] \rangle_t$ gives the average density of calls.

Proposition 3. *The stationary autocorrelation function does not depend on the parameter $L(\in [1, N-1])$, i.e. from the quantity of tasks type A and B, only from the length of the list.*

Proof. It is the consequence of the linearity of the problem. (The details of the proof are shown in the Appendix). □

Time dependence

It is sufficient to study the model with $L = N-1$.

Proposition 4. *The autocorrelation function is a sum of N exponential functions.*

The transition matrix A is (doubly) stochastic which yields that the eigenvalues lie inside the unit circle in the complex plane. The eigenvalue $\lambda_1 = 1$ is non-degenerate because of ergodicity, and the other eigenvalues are $|\lambda_i| < 1$ for $i = 2..N$ For the cases I have studied with concrete form of w_i I found the matrix A to be diagonalisable (numerically). If this is true then

$$A = V^{-1} D V\,, \tag{36}$$

where the rows $\underline{v}_i$ of V are the eigenvectors of A from the left.

$$\underline{P}_A(t) = \sum_{i=1}^{N} a_i \underline{v}_i A^t = \sum_{i=1}^{N} a_i \underline{v}_i \lambda_i^t \tag{37}$$

$$P_A(1,t) = \sum_{i=1}^{N} a_i v_{i,1} \lambda_i^t \tag{38}$$

We need the concrete form of the w_i transition probabilities to get further. The list is assumed to be arranged by priorities, therefore w_i is monotonically decreasing in i. I have studied several examples for w_i, but only the model with power-law decaying w_i can produce a power-law decaying region in the autocorrelation function. For example when w_i is exponential in i, the autocorrelation function decays more slowly than power function with a system size dependent cutoff at the end.

We will study the model with $w_i \sim i^{-\sigma}$, where σ is a parameter of the model.

4.4 Interevent time distribution

Expectation values (waiting times)

Let us investigate the expected timing of execution for the i^{th} element of the list. Let τ_i be the random time when the task with initial position i gets to the first position. We calculate the expectation value for τ_i by conditioning on the step before. Let $E_{1..i-1}$ denote the event that we have chosen from the first $i-1$ positions (then the i^{th} element does not move), E_i be the event of choosing the i^{th} element (then it moves to the front of the list) and $E_{i+1..N}$ be the event of choosing a position greater than i in the last step (then the i^{th} elements steps to the right).

$$\mathbb{E}(\tau_i) = \mathbb{E}(\tau_i|E_{1..i-1})\mathbb{P}(E_{1..i-1}) + \mathbb{E}(\tau_i|E_i)\mathbb{P}(E_i) + \mathbb{E}(\tau_i|E_{i+1..N})\mathbb{P}(E_{i+1..N}) \tag{39}$$

$$= [\mathbb{E}(\tau_i)+1]\sum_{j=1}^{i-1} w_j + w_i + [\mathbb{E}(\tau_{i+1})+1]\sum_{j=i+1}^{N} w_j \tag{40}$$

The waiting time for the i^{th} element can be expressed by the waiting time of the $i+1^{th}$ element:

$$\mathbb{E}(\tau_i) = \frac{1+\mathbb{E}(\tau_{i+1})\sum_{j=i+1}^{N} w_j}{1-\sum_{j=1}^{i-1} w_j} \tag{41}$$

The waiting time for the last element of the list is (optimistic) geometrically distributed with parameter w_N, because in each step the N^{th} element is chosen with probability w_N.

$$\mathbb{P}(\tau_N = k) = (1-w_N)^{k-1} w_N \tag{42}$$

$$\mathbb{E}(\tau_N) = \frac{1}{w_N} \tag{43}$$

The recursion can be solved after knowing this:

$$\mathbb{E}(\tau_k) = \frac{1+N-k}{\sum_{j=k}^{N} w_j} \tag{44}$$

The expected waiting time for the first element of the list is N independently from the transition probabilities w_i (while they guarantee ergodicity). This is a manifestation of Little's law [23].

Recursion for generating functions (waiting times)

With similar train of thought we can find a recursion equation for the distribution of τ_ls (strictly speaking, for their generating function).

$$\mathbb{P}(\tau_l = k) = \mathbb{P}(\tau_l = k|E_{1..l-1})\mathbb{P}(E_{1..l-1}) + \mathbb{P}(\tau_l = k|E_l)\mathbb{P}(E_i) + \mathbb{P}(\tau_l = k|E_{l+1..N})\mathbb{P}(E_{l+1..N}) \tag{45}$$

$$= \mathbb{P}(\tau_l = k-1)\sum_{i=1}^{l-1} w_i + \delta_{k,1} w_l + \mathbb{P}(\tau_{l+1} = k-1)\sum_{i=l+1}^{N} w_i \tag{46}$$

We multiply these equations with z^k and we sum them from $k=1$ to ∞. Let $G_{\tau_l}(z)$ denote the generating function of τ_l:

$$G_{\tau_l}(z) \equiv \sum_{k=1}^{\infty} \mathbb{P}(\tau_l = k) z^k \tag{47}$$

and the recursion equation

$$G_{\tau_l}(z) = \frac{z w_l + z G_{\tau_{l+1}}(z) \sum_{i=l+1}^{N} w_i}{1 - z \sum_{i=1}^{l-1} w_i} \tag{48}$$

we know the generating function for τ_N (from (42) and (47)):

$$G_{\tau_N}(z) = \sum_{k=1}^{\infty} (1-w_N)^{k-1} w_N z^k = z w_N \sum_{k=0}^{\infty} [z(1-w_N)]^k = \frac{z w_N}{1 - z(1-w_N)} \tag{49}$$

This generating function is meaningful for $z < \frac{1}{1-w_N}$. I could not solve the recursion analytically, but I used these formulae to calculate the interevent time distribution by numerical methods.

Remark. *The recursion* (41) *for the expectation values can be derived from the recursion* (48) *for the generating functions by differentiating the latter in* $z = 1$.

In the $L = N-1$ model (#of$A = 1$) the interevent time distribution is the waiting time for the first element of the list. This model is going to be the most interesting because it can produce power-law decaying interevent time distribution. The probability mass function can be calculated by differentiating the generating function:

$$P(\tau_l = k) = \frac{1}{k!} G_{\tau_l}^{(k)}(0) \tag{50}$$

This equation is going to be evaluated numerically in the next subsection.

When $1 < L < N-1$ there are more As in the list, and the interevent time is the waiting time until an A is getting to the first position. If the tasks of type A are in the $i_1 \dots i_{N-L}$ positions and $i_1 = 1$, then the interevent time reads

$$\tau_A = \min_k \{\tau_{i_k}\} \tag{51}$$

We need the joint probability distributions of the random variables τ_{i_k} to compute the distribution of the interevent times. This is a hard problem in general and I am going to analyse the L dependence of the interevent time distribution by simulations.

The $L = 1$ case can be solved analytically. If we know the position of the (unique) task B, then the probability mass function of the interevent time τ_A can be determined. The stationary distribution is uniform, which means that the stationary interevent time distribution has to be calculated such that the position of B is uniform in $2 \dots N$. Applying the law of total probability yields

$$\mathbb{P}(\tau_A = 1) = \sum_{k=2}^{N} \mathbb{P}(\tau_A = 1 | \{pos(B) = k\}) \mathbb{P}(\{pos(B) = k\}) \tag{52}$$

$$= \sum_{k=2}^{N} (1-w_k) \frac{1}{N-1} = 1 - \frac{1-w_1}{N-1} \tag{53}$$

$$\mathbb{P}(\tau_A = l) = \sum_{k=2}^{N} \mathbb{P}(\tau_A = l | \{pos(B) = k\}) \mathbb{P}(\{pos(B) = k\}) \tag{54}$$

$$= \sum_{k=2}^{N} w_k (1-w_1) w_1^{l-2} \frac{1}{N-1} = (1-w_1)^2 w_1^{l-2} \frac{1}{N-1} \tag{55}$$

which is an exponential decay. We used the notation $pos(B)$ for the position of task B.

4.5 Number of bursty events

Bursty periods are sequences of events straddled by pure sequences of B of fixed length (which is defined as the time window). The quantity of tasks type A in a bursty train is denoted by E. A bursty period is demonstrated on an example below, where the time window Δt is chosen to be 4:

$$\ldots BBBBABAABABBAAAAABAABBBAABBBB\ldots$$

The distribution of number of bursty events for this model is the dual of the interevent time distribution in some sense. Let us investigate the limits of the list composition.

Example. $L = N-1$ *case*

The list contains only one task type A, and the interevent times are independent identically distributed. Form equation (2) (we mean the discrete version of that formula) it follows that the $P(E)$ distribution has exponential decay (where the exponent depends on the time window).

Example. $L = 1$ *case*

If the time window is chosen to be 1, then the P_E probability mass function is connected to the waiting time of the first element of the list (which is the only element of type B):

$$P(E = k) = \mathbb{P}(\tau_1 = k+1 | \tau_1 \geq 2) = \frac{\mathbb{P}(\tau_1 = k+1)}{1 - \mathbb{P}(\tau_1 = 1)} \mathbb{I}(k \geq 2)\,, \tag{56}$$

for which we are going to see a power-law decay in the following subsection.

I have run simulations to analyse the L and Δt dependence of the model (fig.20). When $1 <$

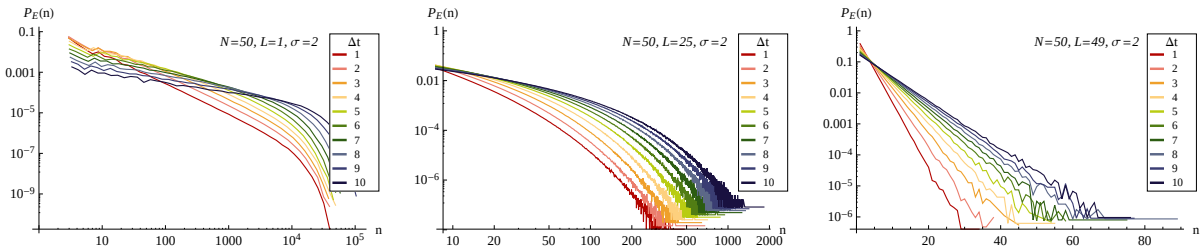


Figure 20: Probability mass function of the number of bursty events (simulation results). The decay gets slower as the time window is increased.

$L < N-1$ we observe a crossover between the power and the exponential behaviour: the decay of the P_E distribution is faster than a power but slower than an exponential function.

4.6 Numerical results

Autocorrelation function

I have done calculations for the stationary autocorrelation function for various values of list sizes (N) and transition probabilities (σ).

$$w_i = \frac{i^{-\sigma}}{\sum_{i=1}^{N} i^{-\sigma}} \tag{57}$$

The probability of choosing an element from the tail of the list is small, that is why we expect the autocorrelation function not to depend on the length of the list on short times (if σ is large enough).
When σ is increased, the probability of choosing an element from the beginning of the list is also increased, and the decay of the autocorrelation function is getting slower.

Program code The program calculates the autocorrelation function by multiplying the initial vector by the transition matrix after one by one. After every 10^{th} steps the timestep is changed to 10 times larger (by using the 10^{th} power of the transition matrix).

Results The results of the computation are shown in fig.21. The autocorrelation functions show

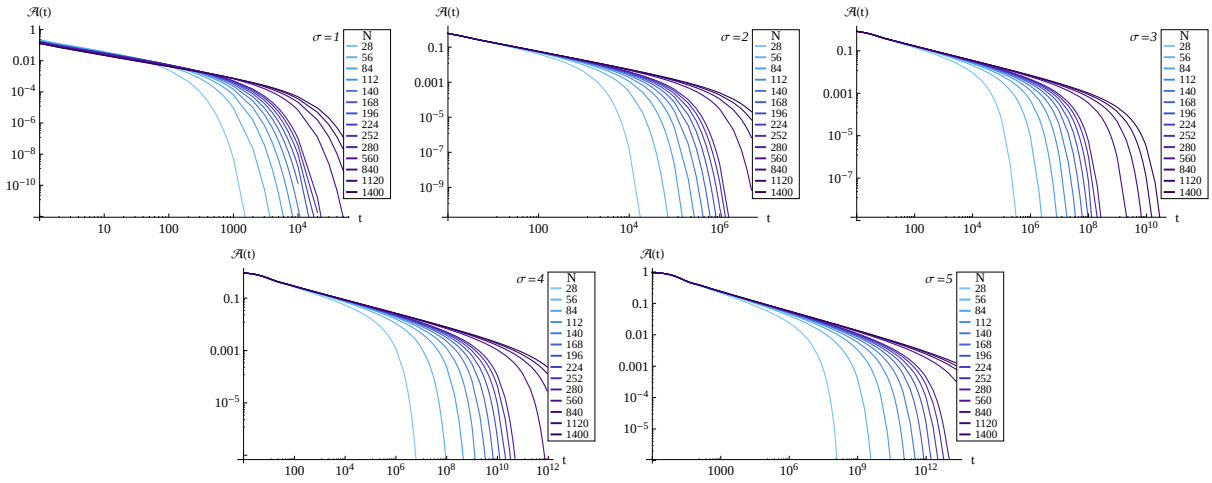


Figure 21: Stationary autocorrelation functions for various list sizes ($\sigma = 1 \ldots 5$)

.

a wide power-law decaying region with an exponential cutoff. This cutoff can be explained by equation (38): for large t the autocorrelation function decays exponentially with the second largest eigenvalue. As the size of the list is increased, the cutoff moves further. This suggests that the stationary autocorrelation function is power-law decaying in the $N \to \infty$ limit. This behaviour is given as a sum of infinitely many exponential functions.

We can also notice that the transient region of the autocorrelation functions fit each other for various list sizes when σ is fixed. This behaviour is determined by the smallest eigenvalues of the transition matrix.

Observation. *Given σ, the autocorrelation functions for various list sizes can be rescaled to fit each other except for the transient region (fig.22, equation* (58)*).*

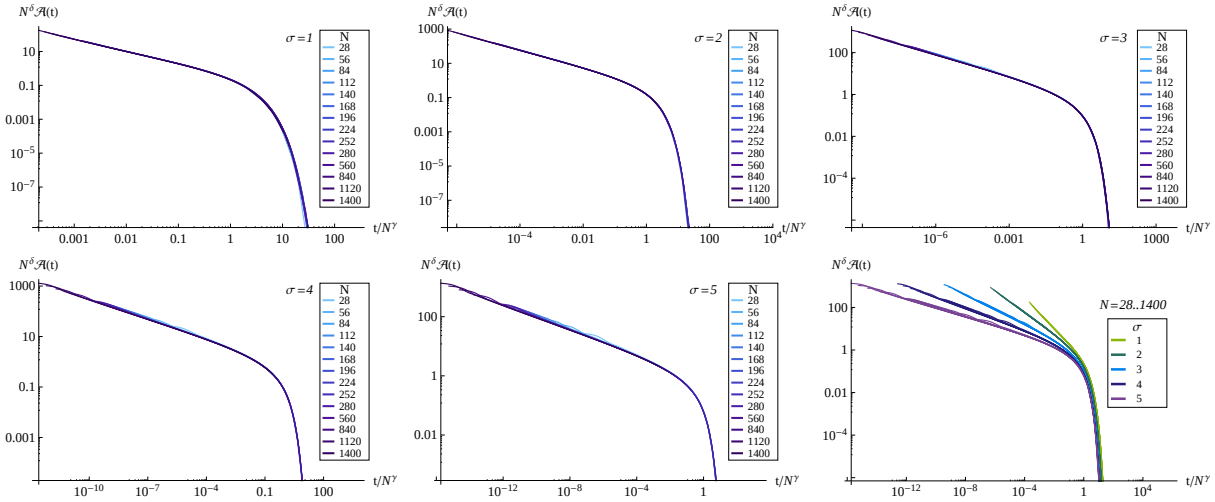


Figure 22: Stationary autocorrelation functions can be rescaled to fit each other when σ is fixed. The scaling parameters γ and δ are listed in tab.3. The curves determined by the scaled functions are described by $f_{\sigma}(t)$ (equation (58))

$$\mathscr{A}(t,N,\sigma) = N^{-\delta} f_{\sigma}\left(\frac{t}{N^{\gamma}}\right) \tag{58}$$

The scaling parameters γ and δ corresponding to different values of σ are listed in tab.3.

σ	0.5	0.8	1	1.5	1.8	2	3	4	5
γ	1.07	1.15	1.17	1.57	1.8	2	3	4	5
δ	1	1	1	1	1	1	1	1	1

Table 3: Scaling parameters γ and δ for various values of σ.

Equation (58) implies that the stationary autocorrelation function is a generalised homogeneous function of the length of the list and of time (the transient region is left out of consideration).

$$\mathscr{A}(b^{y_t}t, \frac{N}{b}, \sigma) = b^{\delta} N^{-\delta} f_{\sigma}\left(\frac{b^{y_t} b^{\gamma} t}{N^{\gamma}}\right), \tag{59}$$

if we substitute $y_t = -\gamma$,

$$\mathscr{A}(t,N,\sigma) = b^{-\delta}\mathscr{A}(b^{-\gamma}t, \frac{N}{b}, \sigma) \tag{60}$$

The scaling function f_{σ} can be expressed from this:

$$f_{\sigma}\left(\frac{t}{N^{\gamma}}\right) = \mathscr{A}(\frac{t}{N^{\gamma}}, 1, \sigma) \tag{61}$$

Equation (58) can be written in a different form (by using $b = t^{\frac{1}{\gamma}}$):

$$\mathscr{A}(t,N,\sigma) = t^{-\frac{\delta}{\gamma}} \tilde{f}_{\sigma}\left(\frac{t}{N^{\gamma}}\right), \tag{62}$$

where

$$\tilde{f}_\sigma\left(\frac{t}{N^\gamma}\right) = \mathscr{A}(1, \frac{N}{t^{1/\gamma}}, \sigma). \tag{63}$$

We conclude that between the transient and the cutoff region the autocorrelation function has a power decay:

$$\mathscr{A}(t) \sim t^{-\alpha}. \tag{64}$$

Equation (62) allows us to express α in terms of γ and δ when $N \to \infty$.

$$\mathscr{A}(t, N, \sigma) = t^{-\frac{\delta}{\gamma}} \tilde{f}_\sigma\left(\frac{t}{N^\gamma}\right) \to t^{-\frac{\delta}{\gamma}} \tilde{f}_\sigma(0)\,, \tag{65}$$

thus $\alpha = \frac{\delta}{\gamma}$. Comparing this with tab.3 gives rise to a conjecture between the model parameter σ and the exponent of the autocorrelation function.

Conjecture 1. *When* $N \to \infty$

$$\alpha = \frac{1}{\sigma}\,, \tag{66}$$

and the difference from it may be caused by finite size effects.

This conjecture is in agreement with the fits of power function to the autocorrelation functions (tab.4). The conjecture is going to be proved for $\sigma = 2$ and demonstrated for other integer values of σ at the end of this section.

σ	1	2	3	4	5
α	0,755	0,535	0,356	0,274	0,222

Table 4: Exponents of the power decay of the autocorrelation function (and of $f_\sigma(t)$).

Interevent time distribution

$L = N - 1$ **model** The calculations were done by numerically evaluating the equation (50) for the first element of the list. The results are shown on fig.23.(a) for various values of σ. The interevent time distribution can be approximated by a power function on a wide range.

$$P(t_{ie}) \sim t_{ie}^{-\beta}\,, \tag{67}$$

which is followed by an exponential cutoff. The exponential decay $\sim (1 - w_N)^t$ comes from the event that the task A has reached the end of the list (which have occurred with high probability when t is large).

I have determined the exponents for the power-decaying region for various values of σ (tab.5). The results suggest that the exponents satisfy the following formula $\frac{2\sigma-1}{\sigma}$.

Conjecture 2.

$$\beta = 2 - \frac{1}{\sigma} \tag{68}$$

We can observe small waves in the interevent time mass function for high values of σ. This is in correspondence with the increasing transient region in the autocorrelation function, which is a consequence of opening a gap at small λ in the transition matrix (30) of the Master equation.

The conjectures for the values of α and β suggest a scaling relation:

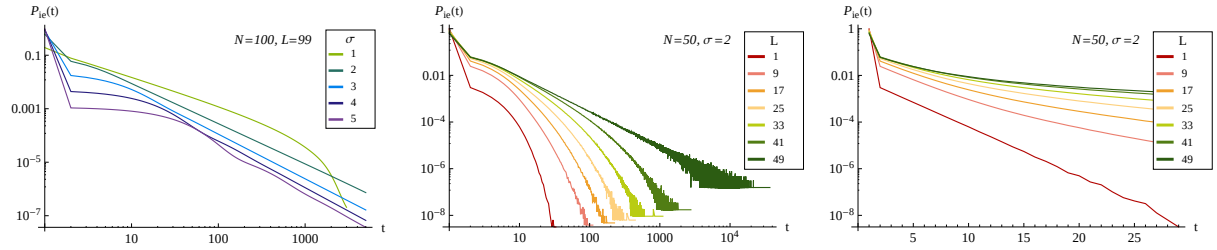
Conjecture 3.

$$\alpha + \beta = 2 \tag{69}$$

Assuming equation (69) we have to calculate only one of the exponents and the other will be determined by the scaling relation. In section 4.7 I am going to calculate the exponent for the interevent time distribution for integer values of σ. At last in section 4.8 I am going to prove conjecture 3 for a wide range of models.

σ	2	3	4	5
β	1.501	1.668	1.753	1.7998

Table 5: Exponents of the power decay of interevent time distributions $P(\tau_1)$ for various values of σ in the $L = N - 1$ model.



(a) σ dependence (b) L dependence (c) L dependence

Figure 23: (a) σ dependence of the interevent time distribution in the $L = N - 1$ model (numerical results). (b,c) L dependence of the interevent time distribution for $\sigma = 2$ (simulation results). The axes are logarithmic on (a,b), and semi-logarithmic on (c).

L dependence The $L = N - 1$ model produces power-decaying interevent time distribution and as it has been mentioned before the $L = 1$ model produces exponential decay. I have run simulations for various values of L, the results are shown in fig.23.(b)-(c). For $1 < L < N - 1$ we observe a decay faster than a power, but slower than an exponential function.

4.7 Analytic calculation for the interevent time exponent

When the list contains only one task of type A, both of the autocorrelation function and the interevent time probability mass function are power-law decaying with an exponential cutoff. As the size of the list is increased the position of cutoff goes further. We expect to get a pure power decay in the $N \to \infty$ limit. It might be easier to determine the exponents in the case of infinitely long list because then the power-law decay is not only a finite region but also the *asymptotic* of the functions.

So we want to calculate the recurrence time τ of the first element of the infinite list. Let us deal with the case when $\tau \neq 1$, i.e. the observed element moves to the second position. We set the origin of the time to this event (the first step was in $t = -1$). Let $q_n(t)$ $(n \geq 2)$ denote the probability of finding the observed element at position n after t timesteps (after the first step) without any recurrences up to time t. The restriction not to recur is going to be important because this makes large jumps to the front of the list *forbidden* for the observed element.

$$q_n(0) = (1 - w_1)\delta_{n,2} \tag{70}$$

$$q_n(t+1) = \begin{cases} P_1 q_2(t) & \text{if } n = 2, \\ P_{n-1} q_n(t) + (1 - P_{n-1}) q_{n-1}(t) & \text{if } n > 2 \end{cases} \tag{71}$$

where $P_n = \sum_{k=1}^{n} w_i$. The latter equation expresses that the observed element steps to the right if the position chosen is greater than its current position. The probability of not to recur until time t is $Q(t) = \sum_{n=2}^{\infty} q_n(t)$. The probability of the (first) recurrence at t is equal to $Q(t) - Q(t+1)$.

$$q_2(t+1) = P_1 q_2(t) \tag{72}$$

$$q_3(t+1) = P_2 q_3(t) + (1 - P_2) q_2(t) \tag{73}$$

$$\vdots$$

The probability of not having recurrence is monotonous decreasing in time (by summing the equations above):

$$Q(t+1) = Q(t) - \sum_{k=2}^{\infty} q_k(t) w_k \tag{74}$$

that is, if the observed element is at the k^{th} position and we choose the same position then a recurrence will occur.

Here we introduce the logarithmic generating function (discrete Laplace-transform):

$$\Gamma(\lambda, t) = \sum_{k=2}^{\infty} q_k(t) e^{-k\lambda} \tag{75}$$

From this quantity we can calculate the required probability $Q(t) = \Gamma(0, t)$. The boundary conditions:

$$\Gamma(\lambda, 0) = (1 - w_1) e^{-2\lambda} \tag{76}$$

$$\Gamma(\lambda \to \infty, t) = (1 - w_1) w_1^t e^{-2\lambda} \,. \tag{77}$$

The second equation is the consequence of the fact that in $\sum_{k=2}^{\infty} q_k(t) e^{-\lambda k}$ the leading order term is when $k = 2$, and $q_2 = (1 - w_1) w_1^t$ (see equations (70)-(71)).
Time evolution:

$$\Gamma(\lambda, t+1) = \sum_{k=2}^{\infty} P_{k-1} q_k(t) e^{-k\lambda} + \sum_{k=2}^{\infty} (1 - P_k) q_k(t) e^{-(k+1)\lambda} \tag{78}$$

$$= \Gamma(\lambda, t) - \sum_{k=2}^{\infty} (1 - P_{k-1}) q_k(t) e^{-k\lambda} + \sum_{k=2}^{\infty} (1 - P_k) q_k(t) e^{-(k+1)\lambda} \tag{79}$$

We are interested in the case when $w_i \sim i^{-\sigma}$ and we use integral approximation for $P(k)$:

Integral approximation

$$P_k \approx \frac{\int_1^{k+1} i^{-\sigma} di}{\int_1^{\infty} i^{-\sigma} di} = 1-(k+1)^{-\sigma+1} \tag{80}$$

$$\Gamma(\lambda,t+1) = \Gamma(\lambda,t) - \sum_{k=2}^{\infty} k^{-\sigma+1} q_k(t) e^{-k\lambda} + \sum_{k=2}^{\infty} (k+1)^{-\sigma+1} q_k(t) e^{-(k+1)\lambda} \tag{81}$$

We restrict ourselves to integer values of $\sigma(\geq 2)$, and we affect with the operator $(-\frac{\partial}{\partial\lambda})^{\sigma-1}$ on both sides of the equation:

$$\left(-\frac{\partial}{\partial\lambda}\right)^{\sigma-1} [\Gamma(\lambda,t+1)-\Gamma(\lambda,t)] = \left(e^{-\lambda}-1\right)\Gamma(\lambda,t) \tag{82}$$

If we redefine our model in such a way that $\tilde{w}_i = i^{-\sigma+1} - (i+1)^{-\sigma+1}$ (for $i \geq 1$) then the asymptotic remains the same: $w_i \sim i^{-\sigma}$ and for that model $P_k = 1-(k+1)^{-\sigma+1}$. I have also calculated the exponents of the interevent time distribution numerically for this model and it gave the same result as for the original model.

Continuous time approximation

We assume that functions $q_i(t)$ are smooth in variable t especially for large values of t. This is supported by the dynamics in equation (71) which says that $q_i(t+1)$ is a weighted average of 0, $q_i(t)$ and $q_{i-1}(t)$. Then we are using the following approximation: $q_i(t+1)-q_i(t) = \frac{dq_i(t)}{dt}$, or equivalently $\Gamma(\lambda,t+1)-\Gamma(\lambda,t) = \frac{\partial\Gamma(\lambda,t)}{\partial t}$.

$$\left(-\frac{\partial}{\partial\lambda}\right)^{\sigma-1} \frac{\partial\Gamma(\lambda,t)}{\partial t} = (e^{-\lambda}-1)\Gamma(\lambda,t) \tag{83}$$

We are interested in the small λ and large t limit of this equation to get the asymptotic behaviour of $Q(t)$. The linearised version of this equation:

$$\left(-\frac{\partial}{\partial\lambda}\right)^{\sigma-1} \frac{\partial\Gamma(\lambda,t)}{\partial t} = -\lambda\Gamma(\lambda,t)\,. \tag{84}$$

Now we are looking for a scaling solution for this equation:

$$\Gamma(\lambda,t) = t^{-\zeta}\phi\left(\frac{\lambda}{t^{\theta}}\right) \tag{85}$$

Substituting this into (84) and rearranging the terms result in

$$-t^{-\zeta-1-\theta\sigma}\left((\theta(\sigma-1)+\zeta)t^{\theta}\phi^{(\sigma-1)}\left(\frac{\lambda}{t^{\theta}}\right)+\lambda\theta\phi^{(\sigma)}\left(\frac{\lambda}{t^{\theta}}\right)\right) = (-1)^{\sigma}\lambda t^{-\zeta}\phi\left(\frac{\lambda}{t^{\theta}}\right) \tag{86}$$

To get a consistent equation we have to choose θ and ζ such that

$$\zeta = -\theta(\sigma-1) \tag{87}$$

$$1 = -\theta\sigma \tag{88}$$

The solution of these equations are $\theta = -\frac{1}{\sigma}$ and $\zeta = 1 - \frac{1}{\sigma}$, and with them the remaining differential equation reads

$$\phi^{(\sigma)}(\mu) = (-1)^{\sigma}\sigma\phi(\mu)\,. \tag{89}$$

The solutions of this equation are

$$\phi_i(\mu) = e^{-\varepsilon_i \sigma^{\frac{1}{\sigma}}\mu} \tag{90}$$

for $i = 1 \ldots \sigma$ where ε_i are the $\sigma^t h$ roots of unity. Functions ϕ_i behave well at zero, thus we have

$$\Gamma(0,t) = t^{\frac{1}{\sigma}-1} \sum_{i=1}^{\sigma} c_i \phi_i(0) \tag{91}$$

By differentiating $P_{ie}(t) \sim t^{-\left(2-\frac{1}{\sigma}\right)}$, so we have proved conjecture 2 for integer values of σ.

Remark. *The continuous time approximation is the solution of the model with continuous time Markov-chain when tasks are executed corresponding to an exponential rate function.*

Remark. *The asymptotic behaviour of the interevent time distribution can be determined without the continuous time approximation using a second Laplace transform in the time variable. This method requires more calculation, thus it is not shown here.*

4.8 Proof of the scaling relation between the exponents

Let us remind of conjecture 3 which connected the exponents of the autocorrelation function and the interevent time distribution for the $L = N - 1$ models.

$$\alpha + \beta = 2 \tag{92}$$

The essential properties of the model for this scaling relation are that the interevent times are independent and power-law decaying. That is, the scaling law can be extended for any point processes with independent power-law decaying interevent times (i.e. for renewal processes with power-law decaying interevent times). The definition of the autocorrelation function is given in (32):

$$\mathscr{A}(t) = \frac{\mathbb{P}(X(t) = 1 | X(0) = 1) - \langle \mathbb{E}[X(t)] \rangle_t}{1 - \langle \mathbb{E}[X(t)] \rangle_t}\,. \tag{93}$$

Theorem 1. *For any renewal processes with power-law decaying interevent times:*

$$P_{ie}(t) \sim t^{-\beta} \qquad \text{as } t \to \infty,$$

with $\beta \in (1,2)$ the autocorrelation function is power-law decaying

$$\mathscr{A}(t) \sim t^{-\alpha} \qquad \text{as } t \to \infty$$

and the exponents satisfy the scaling relation $\alpha + \beta = 2$.

Proof. Let τ_m denote the recurrence times: $\tau_0 = 0$, $\tau_{m+1} = inf\{t > \tau_m | X(t) = 0\}$, and let T denote the set of recurrence times: $T = \{\tau_m : m = 0, 1, 2, \dots\}$. Thus $X(t)$ is non-zero only for ts that are elements of T. We are going to use the notation $\tau_1 = \tau$ for the interevent times which are independent and identically distributed.
The average density of calls can be given by $\langle \mathbb{E}[X(t)] \rangle_t = \frac{1}{\mathbb{E}[\tau]}$. Using these notations the autocorrelation function can be written in the form

$$\mathscr{A}(t) = \frac{\mathbb{P}(t \in T) - \frac{1}{\mathbb{E}[\tau]}}{1 - \frac{1}{\mathbb{E}[\tau]}} \tag{94}$$

In the case of $\beta \in (1,2)$ the expectation value for the interevent time is infinite, which yields

$$\mathscr{A}(t) = \mathbb{P}(t \in T) \tag{95}$$

Proposition 5. *The Laplace transform of the autocorrelation function can be expressed by the Laplace transform of the interevent time distribution.*

Proof.

$$g(\lambda) \equiv \sum_{t=0}^{\infty} \mathscr{A}(t) e^{-\lambda t} \tag{96}$$

$$\begin{aligned} g(\lambda) &= \sum_{t=0}^{\infty} e^{-\lambda t} \mathbb{P}(t \in T) = \sum_{t=0}^{\infty} e^{-\lambda t} \sum_{m=0}^{\infty} \mathbb{P}(\tau_m = t) = \sum_{m=0}^{\infty} \mathbb{E}\left(e^{-\lambda \tau_m}\right) \\ &= \sum_{m=0}^{\infty} \mathbb{E}\left(e^{-\lambda m \tau}\right) = \left(1 - \mathbb{E}\left[e^{-\lambda \tau}\right]\right)^{-1}, \end{aligned} \tag{97}$$

because τ_m can be written as a sum of m independent τ distributed random variables. □

Definition 1. *A function u is ultimately monotone if $\exists x_0 > 0$: u is monotone in the interval (x_0, ∞).*

Theorem 2. *[Tauber,[24]] Let $0 \leq \rho < \infty$. If u is ultimately monotone with Laplace transform ω and L is slowly varying in infinity then the following statements are equivalent:*

$$u(x) \sim \frac{1}{\Gamma(\rho)} x^{\rho - 1} L(x) \qquad t \to \infty \tag{98}$$

$$\int_0^{\infty} e^{-\lambda x} u(x) \mathrm{d}x \equiv \omega(\lambda) \sim \lambda^{-\rho} L\left(\frac{1}{\lambda}\right) \qquad \lambda \to 0 \tag{99}$$

Direction (98)⇒(99) is called Abelian theorem while the opposite direction is called Tauberian theorem.

We want to use the Abelian theorem for the interevent time distribution. The probability density (or mass) function have to be normalised, which means that $\beta \geq 1$. This corresponds with the domain $\rho \leq 0$ for which the original version of the theorem cannot be applied.

Theorem 3 (Abel, extended version)**.** *Let* $-1 < \rho < 0$*. If u is ultimately monotone with Laplace transform* ω *and* $xu(x)$ *is also ultimately monotone and L is slowly varying in infinity then the asymptotic behaviour of u at infinity and its Laplace transform at* 0 *satisfy the relation:* (100)⇒(101)

$$u(x) \sim \frac{1}{\Gamma(\rho)} x^{\rho-1} L(x) \qquad t \to \infty \tag{100}$$

$$\omega(\lambda) - \int_0^\infty u(x)\mathrm{d}x \sim \lambda^{-\rho} L\left(\frac{1}{\lambda}\right) \qquad \lambda \to 0 \tag{101}$$

Proof. We trace back this statement to the original version of the Tauber theorem. Let $-1 < \rho < 0$, then $v(x) = \frac{1}{\rho} x u(x)$ satisfy the conditions for theorem 2. Let η denote the Laplace transform of v.

$$\int_0^\infty e^{-\lambda x} u(x)\mathrm{d}x = \int_0^\infty \frac{e^{-\lambda x}}{x} \rho v(x)\mathrm{d}x = \int_0^\infty \int_\lambda^\infty e^{-\nu x}\mathrm{d}\nu \rho v(x)\mathrm{d}x = \tag{102}$$

$$= \int_\lambda^\infty \int_0^\infty e^{-\nu x} \rho v(x)\mathrm{d}x\mathrm{d}\nu = \rho \int_\lambda^\infty \eta(\nu)\mathrm{d}\nu = \tag{103}$$

$$= \rho \int_0^\infty \eta(\nu)\mathrm{d}\nu - \rho \int_0^\lambda \eta(\nu)\mathrm{d}\nu = \int_0^\infty u(x)\mathrm{d}x + \lambda^{-\rho} L(\frac{1}{\lambda}) \tag{104}$$

in leading order in λ. □

Now we are ready to evaluate the asymptotic behaviour of the Laplace transforms in (97). Applying the Abelian theorem (100)⇒(101) to the right side of (97) yields

$$g(\lambda) = \left(1 - \mathbb{E}\left[e^{-\lambda\tau}\right]\right)^{-1} \sim \lambda^{1-\beta} \qquad \lambda \to 0 \tag{105}$$

The asymptotic decay of the autocorrelation function can be determined by using Tauberian theorem (99)⇒(98) on $g(\lambda)$:

$$\mathscr{A}(t) \sim t^{\beta-2} \tag{106}$$

□

With the same train of thought we can extend our results to the $\beta \in (2,3)$ region. The result is $\mathscr{A}(t) \sim t^{-(\beta-2)}$ in agreement with [25]. The authors of [25] have not discussed the case when the expectation value is infinite for the interevent times ($\beta \in (1,2)$.

It was essential in the scaling theorem that the interevent times were independent. If the interevent times show long-range dependencies then the autocorrelation exponent may change. In the following subsection we compose a dependent set of interevent times by the Metropolis algorithm and show that the scaling law can be violated when the interevent times are not independent.

Dependent set of power-law decaying interevent times

We use the Metropolis algorithm to obtain a dependent sequence of power-law distributed random variables. For this sake we construct a Markov chain on the integers that has power-law decaying stationary distribution $P(x) \sim x^{-\beta}$.The value of the random variable in the next step depends

only on its current value. The algorithm uses a proposal density (transition rate) $Q(x',x_n)$ which generates a proposal sample. This sample is accepted for the next value with probability

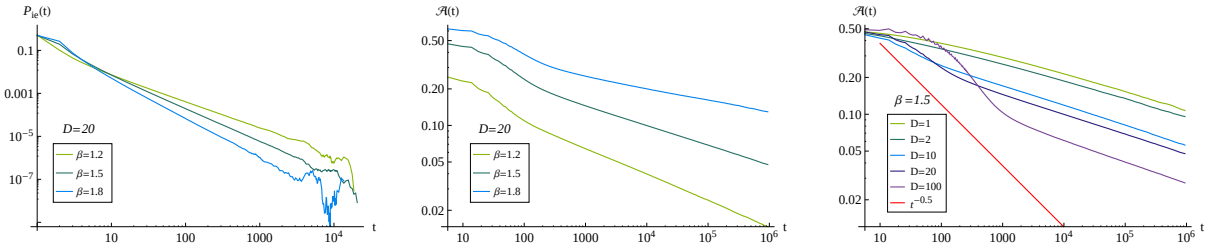
$$\alpha(x',x_n) = \min\left\{\frac{Q(x_n,x')P(x')}{Q(x',x_n)P(x_n)},1\right\} \tag{107}$$

The algorithm:

1. Choose an x_0 as initial value
2. Generate x' from the distribution $Q(x',x_n)$
3. Accept this value as x_{n+1} with probability $\alpha(x',x_n)$, otherwise $x_{n+1} = x_n$
4. Go to step 2

Our goal is to generate a power-law distributed sample with long-range dependency. Therefore the mixing of the Markov chain should be slow, i.e. the gap in the Markov chain should be equal to 0.

In the proposal density we allow only small differences between consecutive interevent times: $Q(x',x_n) = \frac{1}{2D}\mathbb{I}(x_n - D, x_n + D)$ or equivalently x' as a random variable is discrete uniformly distributed: $x' \sim DU[x_n - D, x_n + D]$. The simulation results show that the autocorrelation function decays more slowly than it would be assumed from the scaling law (24).



(a) Interevent time distribution (b) Autocorrelation function (c) Autocorrelation function

Figure 24: Simulation results for the Metropolis algorithm. The exponent of the interevent time distribution is varied in subfigure (a) and (b). In subfigure (c) $\beta = 1.5$ and the parameter D of the proposal density is varied.

Analysing processes with interevent times given from a Markov chain might be useful also for the modelling because this is one of the simplest ways to generate dependent interevent times. However in this section we use the Markov chain of interevent times only to show that the scaling law $\alpha + \beta = 2$ is true for independent processes and can be violated when the interevent times are dependent. In this sense *power-law decaying autocorrelation function signs long-range dependencies only if the exponent violates the scaling relation* (69).

I have derived a formula for the Laplace transform of the autocorrelation function when the interevent times are given from a Markov chain (for details see the Appendix), but I couldn't determine the exponent of the autocorrelation function analytically for the Metropolis algorithm.

4.9 Modified model

I have also studied a modified version of the priority arranged list model. In this model at each step we choose an activity from the list and the chosen activity switches places with its left neighbour if it exists. Besides this the first element of the list is executed in each step. This model has similar properties to the original one. The model defines a Markov chain and the stationary distribution is uniform. The autocorrelation function can be derived from the Master equation and it also does not depend on the quantity of phoning tasks. I studied the $w_i \sim i^{-\sigma}$ function for choosing the i^{th} element of the list. I found numerically that the autocorrelation functions show a power-law regime, and the curves for different list sizes can be rescaled to fit each other. The conjecture for autocorrelation exponent is $\alpha = \frac{1}{\sigma+2}$. The $L = N - 1$ model produces a power-law decaying interevent time distribution with exponent conjectured to be $2 - \frac{1}{\sigma+2}$. The scaling law (69) also holds for this model because the interevent times are independent and numerically power-law decaying. This model has interesting mathematics because the paths of the elements of the list define a nearest neighbour random walk with drift depending on the position. I have calculated the distribution of the maximal distance ξ reached by the random walker before recurrence to the front of the list: $\mathbb{P}(\xi = x) \sim x^{-(\sigma+2)}$. I havent found the connection between this and the autocorrelation exponent. The details of this model can be found in Hungarian in my essay for the Students' Scientific Association contest [26].

5 Summary

I have been studying the human communication behaviour in my MSc thesis. My aim was to find an appropriate model that can reproduce the most essential properties of human communication patterns. These patterns are very complicated and I also had to decide what should be taken as essential property. The most studied quantity in the literature is the interevent time distribution which is found to be power-law decaying for a lot of activities like email, surface mail communication, web-browsing or telecommunication. In many cases the measurements show an average over the users, and I was interested if we can see more structures on the individual level. The power-law decay is usually explained by task-queuing models, however there is another approach that explains the heavy tail in the interevent time distribution with the circadian and weekly cycles. The model they proposed is a cascading nonhomogeneous Poisson process. In my thesis I also wanted to decide between these two approaches. Beside the interevent time distribution, the correlation or dependence between the calls is a similarly important question, but this is barely discussed in literature. I have been doing my work in collaboration with a research group at Aalto University. They have also measured the autocorrelation function and the distribution of the number of bursty events averaged over the dataset. The latter is a quantity introduced by the research group to measure deeper dependencies in time-series.

To make measurements and observations on the individual level, I have got access to a small database which contains the interevent times of 70 individuals over 180 days. I have measured the interevent time distribution, the autocorrelation function, the distribution of the number of burtsy events and the number of daily calls. I have found peaks in the interevent time distribution at about 12 hours for the active and at integer times a day for the less active users. The autocorrelation function is periodic in case of times larger than one day for almost all users in the dataset. I have shown that these peaks and the shape of the long time autocorrelation function can be well described by the circadian pattern, especially by a nonhomogeneous Poisson process. However I have also found that the approximately power-law decaying regime of the interevent time distribution and the autocorrelation function cannot be explained by a cascading Poisson process if the cascade process is also Poissonian. To reproduce the short time behaviour of the measured quantities it is worth studying the processes with slowly decaying autocorrelation function and interevent time distribution, for example task-queuing models. There are two ways to get a good description of the long time behaviour as well: these processes can be considered as cascades triggered by a nonhomogeneous Poisson process or a rate function can be introduced directly to these processes. Analysing these combinations is planned to be done in the future. I have also found that the users tend to change their average activity on the scale of 180 days and this should be considered in an appropriate modelling.

As another part of my work I have studied a task-queuing model that can be thought of as a generalisation of the Barabási-model. I have proved that the autocorrelation function does not depend on the composition of the list. By numerical calculations I found out that the autocorrelation function is power-law decaying with an exponential cutoff. I observed that the curves corresponding to different list sizes can be rescaled to fit each other, and I have determined the exponent of the power-law decay as a function of the model parameters using this property. I have investigated the interevent time and the bursty event number distribution by numerical methods. I have found that these quantities are complementary because if one is power-law decaying then the other is exponential and vice versa. I have analysed the case when the interevent time distribution is power-law

decaying and I have calculated the exponent analytically (for integer values of the model parameter σ). I have noticed that the sum of the autocorrelation and interevent time exponents equals to 2, and I have proved this scaling law for every renewal processes with power-law decaying interevent times. With this theorem and with the calculation for the interevent time exponent I have also determined the asymptotic behaviour of the autocorrelation function.

As a continuation of my work some of my measurements should be repeated in the large telecommunication dataset at Aalto University. I would like to determine the interevent time exponent of the task-queuing model for arbitrary values of the model parameter σ. I am also planning to analyse the effects of introducing a rate function in task-queuing models.

Acknowledgements

I would like to express my thanks to my supervisor, Prof. János Kertész, for guiding me and for creating the good circumstances for my research, and also for the opportunity to do my summer practise at Aalto University. I thank Prof. Bálint Tóth for the consultations and for his smart and useful ideas that have helped me to make an advance and progress in my work in many cases. I also thank Márton Karsai for submitting his measurements and Hang-Hyun Jo for the thought-provoking discussions. Finally I would like to thank my parents, my brother and my girlfriend for their support.

6 Appendix

6.1 Independence of the autocorrelation function from parameter *L*

The autocorrelation function is given in (35) and the initial condition for the stationary autocorrelation function is given in (34). We exploit the linearity of the Master equation, and we write the state vector as a sum of two terms:

$$\underline{P}_A^{(L)}(t_0) = \frac{L}{N-1}(1,0,0,\ldots,0) + \frac{N-L-1}{N-1}(1,1,1,\ldots,1)$$

$$\underline{P}_A^{(L)}(t) = \frac{L}{N-1}(1,0,0,\ldots,0)A^t + \frac{N-L-1}{N-1}(1,1,1,\ldots,1)A^t \tag{108}$$

$$= \frac{L}{N-1}(1,0,0,\ldots,0)A^t + \frac{N-L-1}{N-1}(1,1,1,\ldots,1) \tag{109}$$

For any L the time evolution of the first element of the list can be expressed by the $L = N-1$ model (in which there is only one A in the list).

$$P_A^{(L)}(1,t) = \frac{L}{N-1}P_A^{(N-1)}(1,t) + \frac{N-L-1}{N-1} \tag{110}$$

$$= \frac{L}{N-1}\left[A^t\right]_{11} + \frac{N-L-1}{N-1} \tag{111}$$

From this

$$\mathscr{A}^{(L)}(t) = \frac{P_A^{(L)}(1,t) - \frac{N-L}{N}}{1-\frac{N-L}{N}} = \frac{\frac{L}{N-1}P_A^{(N-1)}(1,t) + \frac{N-L-1}{N-1} - \frac{N-L}{N}}{1-\frac{N-L}{N}} \tag{112}$$

$$= \frac{\frac{L}{N-1}\left(P_A^{(N-1)}(1,t) - \frac{1}{N}\right)}{\frac{L}{N-1}\left(1-\frac{1}{N}\right)} = \frac{P_A^{(N-1)}(1,t) - \frac{1}{N}}{1-\frac{1}{N}} = A(t) \tag{113}$$

The autocorrelation function can be expressed by the left upper element of the powers of the transition matrix.

$$\mathscr{A}(t) = \frac{\left[A^t\right]_{11} - \frac{1}{N}}{1-\frac{1}{N}} \tag{114}$$

6.2 Autocorrelation function for interevent times given from a Markov chain

I have derived a formula for the Laplace transform of the autocorrelation function when the interevent times are given from a Markov chain. Let B denote the transition matrix of the Markov chain and let us introduce the diagonal matrix $D(\lambda)$ whose matrix elements are $D_{s,t}(\lambda) = \delta_{st}e^{-\lambda s}$. The Laplace transform of the autocorrelation function can be expressed by the resolvent of the matrix $BD(\lambda)$:

$$g(\lambda) = 1 + \sum_{m,n}^{\infty} \mathbb{P}(\tau_1 = m)\left(1 - D(\lambda)B\right)^{-1}_{mn} e^{-\lambda n} \tag{115}$$

Proof. The description will be similar to the case of independent variables. Let T_m denote the recurrence times: $T_0 = 0$, $T_{m+1} = inf\{t > T_m | X(t) = 0\}$ and T denote the set of recurrence times: $T = \{T_m : m = 0, 1, 2, \dots\}$. The interevent times: $\tau_i = T_i - T_{i-1}$, which are generated from a Markov chain. We use the following autocorrelation function $A(t) = \mathbb{P}(t \in T)$ from equation (95), and we calculate the Laplace transform similarly to the independent case:

$$g(\lambda) \equiv \sum_{n=0}^{\infty} A(n) e^{-\lambda n} = \sum_{m=0}^{\infty} \mathbb{E}\left[e^{-\lambda T_m}\right] = \sum_{m=0}^{\infty} \mathbb{E}\left[e^{-\lambda \sum_{i=1}^{m} \tau_i}\right] \tag{116}$$

By the definition of the transition matrix of the Markov chain:

$$\mathbb{P}(\tau_{i+1} = s) = \sum_{k} \mathbb{P}(\tau_i = k) B_{k,s} \tag{117}$$

We are interested in the stationary autocorrelation function, i.e. $\mathbb{P}(\tau_1 = k) \sim k^{-\beta}$.

$$\mathbb{E}\left[e^{-\lambda T_0}\right] = 1 \tag{118}$$

$$\mathbb{E}\left[e^{-\lambda T_1}\right] = \mathbb{E}\left[e^{-\lambda \tau_1}\right] = \sum_{n=0}^{\infty} \mathbb{P}(\tau_1 = n) e^{-\lambda n} \tag{119}$$

$$\mathbb{E}\left[e^{-\lambda T_2}\right] = \mathbb{E}\left[e^{-\lambda(\tau_1+\tau_2)}\right] = \mathbb{E}\left[\mathbb{E}\left[e^{-\lambda \tau_1} e^{-\lambda \tau_2} \mid \tau_1\right]\right] \tag{120}$$

$$= \mathbb{E}\left[e^{-\lambda \tau_1} \mathbb{E}\left[e^{-\lambda \tau_2} \mid \tau_1\right]\right] \tag{121}$$

To continue we have to calculate

$$\mathbb{E}\left[e^{-\lambda \tau_2} \mid \tau_1\right] = \sum_{n=0}^{\infty} \mathbb{P}(\tau_2 = n \,|\, \tau_1) e^{-\lambda n} = \sum_{n=0}^{\infty} B_{\tau_1 n} e^{-\lambda n} \tag{122}$$

With this

$$\mathbb{E}\left[e^{-\lambda T_2}\right] = \mathbb{E}\left[e^{-\lambda \tau_1} \sum_{n=0}^{\infty} B_{\tau_1 n} e^{-\lambda n}\right] \tag{123}$$

$$= \sum_{m,n=0}^{\infty} \mathbb{P}(\tau_1 = m) e^{-\lambda m} B_{mn} e^{-\lambda n} \tag{124}$$

and similarly

$$\mathbb{E}\left[e^{-\lambda T_3}\right] = \sum_{m,n,k=0}^{\infty} \mathbb{P}(\tau_1 = m) e^{-\lambda m} B_{mn} e^{-\lambda n} B_{nl} e^{-\lambda l} \tag{125}$$

$$\vdots$$

Let us introduce the matrix $D(\lambda)$ whose matrix elements are $D_{s,t}(\lambda) = \delta_{st} e^{-\lambda s}$. With this quantity we can write the general term $\mathbb{E}\left[e^{-\lambda T_k}\right]$ for $k \geq 1$:

$$\mathbb{E}\left[e^{-\lambda T_k}\right] = \sum_{m,n}^{\infty} \mathbb{P}(\tau_1 = m) \left(D(\lambda) B\right)^{k-1}_{mn} e^{-\lambda n} \tag{126}$$

Now we can substitute this to $g(\lambda)$:

$$g(\lambda) = \sum_{m=0}^{\infty} \mathbb{E}\left[e^{-\lambda T_m}\right] = 1 + \sum_{k=0}^{\infty} \sum_{m,n}^{\infty} \mathbb{P}(\tau_1 = m) \left(D(\lambda)B\right)^k_{mn} e^{-\lambda n} \tag{127}$$

$$g(\lambda) = 1 + \sum_{m,n}^{\infty} \mathbb{P}(\tau_1 = m) \left(1 - D(\lambda)B\right)^{-1}_{mn} e^{-\lambda n} \tag{128}$$

If every row of B is the same, i.e. the interevent times are independent, then this formula reduces to $g(\lambda) = \left(1 - \mathbb{E}\left[e^{-\lambda \tau_1}\right]\right)^{-1}$. □

We are interested in the $\lambda \to 0$ asymptotic of this formula, but either this may be hard to calculate for a concrete transition matrix B.

References

[1] David Lazer et al., *"Computational Social Science"*, Science, (2009) pp. 721-723

[2] J.-P. Eckmann, E. Moses, D. Sergi, *"Entropy of dialogues creates coherent structures in e-mail traffic"*, Proc. Natl Acd. Sci. USA **101**, 14333 (2004).

[3] H. Ebel, L.-I. Mielsch, and S. Bornholdt, *"Scale-free topology of e-mail networks"*, Phys. Rev. E **66**, R35103 (2002).

[4] W. Aiello, F. Chung and L. Lu *"Proceedings of the 32nd ACM Symposium on the Theory of Computing"*, (ACM, New York), pp. 171-180. (2000)

[5] T. Zhou, L.-L. Jiang, R.-Q. Su and Y. C. Zhang, *"Effect of initial configuration on network-based recommendation"*, Euro. Phys. Lett. **81**, 58004 (2008).

[6] A. Vázquez et al., *"Modeling bursts and heavy tails in human dynamics"*, Phys. Rev. E, **73**, 036127 (2006).

[7] M. Karsai et al., *"Small but slow world: How network topology and burstiness slow down spreading"*, Phys. Rev. E **83** 025102 (2011)

[8] J. G. Oliveira and A.-L. Barabási, *"Darwin and Einstein correspondence patterns"*, Nature (London) **437**, 1251 (2005).

[9] A.-L. Barabási, *"The origin of bursts and heavy tails in humans dynamics "*, Nature (London) **435**, 207 (2005)

[10] A. Vazquez, *"Exact Results for the Barabási Model of Human Dynamics"*, Phys. Rev. Lett. **95**, 248701 (2005).

[11] W. Hong, X.P. Han, T. Zhou, B.H. Wang *"Heavy-tailed statistics in short-message communication"*, Chinese Physics Letters **26** 028902 (2009)

[12] R.D. Malmgren, D.B. Stouffer, A.E. Motter, L.A.N. Amaral, *"A Poissonian explanation for heavy tails in e-mail communication"*, Proc. Nat. Acad. Sci. **105**, 18153 (2008).

[13] C. Anteneodo, R.D. Malmgren and D.R. Chialvo, *"Poissonian bursts in e-mail correspondence"*, Eur. Phys. J. B: Condens. Matter Complex Syst. **75** 389-94 (2010)

[14] H-H. Jo, M. Karsai, J. Kertész, K. Kaski, *"Circadian pattern and burstiness in mobile phone communication"*, New Journal of Physics **14**, 013055 (2012).

[15] Y. Wu, C. Zhou, J. Xiao, J. Kurths and H.J. Schellnhuber *"Evidence for a bimodal distribution in human communication"*, Proc. Natl Acad. Sci. USA **107** 18803-8 (2010)

[16] X.P. Han, T. Zhou and B.H. Wang, *"Human dynamics with adaptive interest"*, New J. Phys.**10**, 073010. (2008)

[17] T. Zhou, X.P. Han, B.H. Wang *"Towards the understanding of human dynamics"*, arXiv: 0801.1389v1 (2008)

[18] J.-P. Onnela, et al., *"Structure and tie strengths in mobile communication networks"*, Proc. Nat. Acad. Sci. **104**, 7332-7336 (2007)

[19] J.-P. Onnela, et al. *"Analysis of a large-scale weighted network of one-to-one human communication"*, New J. Phys. **9**, 179 (2007)

[20] M. Karsai, K. Kaski, A.-L. Barabási, J. Kertész *"Universal features of correlated bursty behaviour"*, Scientific Reports (Nature) **2**, 397 (2012)

[21] Á. Corral, *"Long-Term Clustering, Scaling, and Universality in the Temporal Occurrence of Earthquakes"*, Phys. Rev. Lett. **92**, 108501 (2004).

[22] T. Kemuriyama, et al., *"A power-law distribution of inter-spike intervals in renal sympathetic nerve activity in salt-sensitive hypertension-induced chronic heart failure."*, BioSystems **101**, 144-147 (2010).

[23] J.D.C. Little, *"A proof of the Queuing formula: $L = \lambda W$"*, Operations Research **9**, 3 383-387 (1961)

[24] W. Feller, *"An Introduction to Probability Theory and its Applications"* 2nd ed. Vol. 2., Wiley India Pvt. Ltd. (2008).

[25] Lipsky, Fiorini *"Auto-correlation of counting processes associated with renewal processes"*, Technical report, Connecticut (1995).

[26] Sz. Vajna, *"Modelling of communication dynamics"* (in hungarian), essay for the Students' Scientific Association contest, Budapest University of Technology and Economics (2011)

[27] M. Abramowitz; I.A. Stegun, eds. (1972), *"Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables"*, New York: Dover Publications, ISBN 978-0-486-61272-0