On The Waiting Time for A Non-Markovian M/M/1 Queueing System

Xian-Yuan Wu
e-mail:wuxy@cnu.edu.cn

School of Mathematical Sciences, Capital Normal University, Beijing, 100048, P. R. China

Xuzhou Jiangsu, July 15, 2016

1 Correspondence patterns of Darwin and Einstein

In 2005, Oliveira and Barabási [Nature, 437, p. 1251] reported their research results on the correspondence patterns of Darwin and Einstein: during their lifetimes, Darwin and Einstein answered a fraction of letters they received (the over-all response rate being 0.32 and 0.24, respectively), and the distributions of response times to letters are both well approximated with a power-law tail that has an exponent $\alpha = 3/2$.

BRIEF COMMUNICATIONS

Darwin and Einstein correspondence patterns

These scientists prioritized their replies to letters in the same way that people rate their e-mails today.

In an era when letters were the main means of exchanging scientific ideas and results, Charles Darwin (1809–82) and Albert Einstein (1879–1955) were notably prolific correspondents. But did their patterns of communication differ from those associated with the instant-access e-mail of modern times? Here we show that, although the means have changed, the communication dynamics have not: Darwin's and Einstein's patterns of correspondence and today's electronic exchanges follow the same scaling laws. However, the response times of their surface-mail communication is described by a different scaling exponent from e-mail communication, providing evidence for a new class of phenomena in human dynamics.

During their lifetimes, Darwin sent at least 7,591 letters and received 6,530; Einstein sent more than 14,500 and received more than 16,200. We start from a record containing the sender, recipient and the date of each letter[1,2] sent or received by the two scientists. Their correspondence exploded after their rise to fame, and reached a highly fluctuating pattern afterwards (Fig. 1a). Although, on average, they wrote 0.59 (Darwin) and 1.02 (Einstein) letters a day during the last 30 years of their lives, these averages hide significant daily fluctuations. For example, Darwin wrote 12 letters on New Year's Day in 1874 and Einstein received 120 letters on 14 March 1949, his 70th birthday.

The response time, τ, represents the time interval between the date a letter was received and the date that the reply was sent. As shown in Fig. 1b,c, the probability that a letter will be replied to in τ days is well approximated by a power law, $P(\tau) \approx \tau^{-\alpha}$, where $\alpha = 3/2$. The fact that the scaling spans close to four orders of magnitude, from days to years, indicates that most responses (53% for Einstein, 63% for Darwin) were sent within less than ten days.

In some cases, however, the correspondence was stalled for months or years. Some of these represent long breaks in the correspondence and a few are a consequence of missing letters. Others, however, correspond to genuine delays, like Einstein's response on 14 October 1921 to Ralph De Laer Kronig's letter of 26 September 1920, which starts with: "In the course of eating myself through a mountain of correspondence I find your interesting letter from September of last year."

To understand the origin of the observed scaling behaviour, we have to realize that, given the wide range of response times, both Darwin and Einstein must have prioritized correspondence in need of a response. Thus, a simple model of their correspondence assumes that letters arrive at a rate λ and are answered at a rate μ. Each letter is assigned a priority, with high-priority letters being answered soon after their arrival, and others having to wait.

The waiting-time distribution of this simple model[3] follows[4] $P(\tau) \approx \tau^{-3/2}\exp(-\tau/\tau_0)$, which predicts a power-law waiting time for the critical regime $\lambda = \mu$, when $\tau_0 = \infty$. Given that Darwin and Einstein answered only a fraction of letters they received (their overall response rate being 0.32 and 0.24, respectively), we have $\lambda > \mu$. This places the model in the supercritical regime, where a finite fraction of letters are never answered. Numerical simulations (see supplementary information) indicate that in this supercritical regime the waiting-time distribution of the responded letters also follows a power law with exponent $\alpha = 3/2$, which is different from the $\alpha = 1$ obtained for e-mail communications[5]. Therefore, although the response times in e-mail and mail communications follow the same scaling law, they belong to different universality classes.

The correspondence patterns of Einstein and Darwin are examples of well mapped patterns of human interaction, but are also of historical interest. Their timely responses to most letters show that they were both aware of the importance of this intellectual intercourse. Occasional delays were not always without consequence. For example, on 14 October 1921 Einstein returned to a correspondence with Theodor Kaluza that he had left off two years earlier, when he discouraged Kaluza from publishing one of his papers: having second thoughts, he recommended that the paper be submitted. Encouraged by this, Kaluza published his famous paper on five-dimensional unified field theory[6], a key component of today's string theory. Would it have changed the course of science if Einstein had not wavered for two years? We shall never know. But our results indicate that Darwin's and Einstein's late responses or resumed correspondences are not singularities or exceptions: they are part of a universal scaling law[7], representing a fundamental pattern of human dynamics that the famous are no better at escaping than the less distinguished.

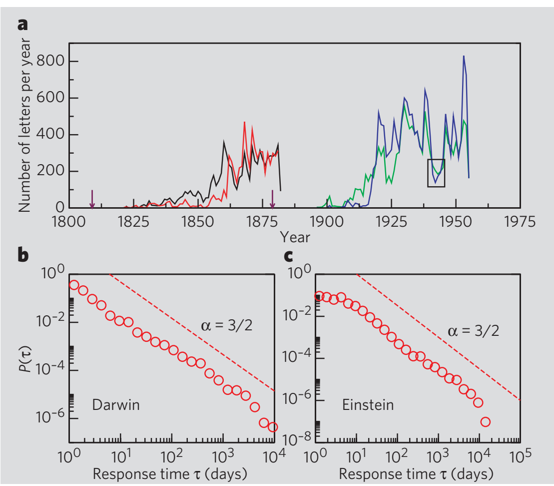

Figure 1 | The correspondence patterns of Darwin and Einstein. **a,** Historical record of the number of letters sent (Darwin, black; Einstein, green) and received (Darwin, red; Einstein, blue) each year by the two scientists[1,2]. An anomalous drop in Einstein's correspondence marks the Second World War period (1939–45, boxed). Arrows, birth dates of Darwin (left) and Einstein (right). **b, c,** Distribution of response times to letters by Darwin and Einstein, respectively. Note that both distributions are well approximated with a power-law tail that has an exponent $\alpha = 3/2$, the best fit over the whole data for Darwin giving $\alpha = 1.45 \pm 0.1$ and for Einstein $\alpha = 1.47 \pm 0.1$.

João Gama Oliveira*†, Albert-László Barabási*‡
*Center for Complex Network Research and Department of Physics, University of Notre Dame, Indiana 46556, USA
e-mail: alb@nd.edu
†Departamento de Física, Universidade de Aveiro, 3810-193 Aveiro, Portugal
‡Center for Cancer Systems Biology, Dana-Farber Cancer Institute, Harvard University, Boston, Massachusetts 02115, USA

1. *The Correspondence of Charles Darwin* Vols 1–14 (Cambridge Univ. Press, Cambridge, 1984–2004).
2. The Collected Papers of Albert Einstein Vols 1,5,8,9 (Princeton Univ. Press, New Jersey, 1993–2004).
3. Cobham, A. *J. Op. Res. Soc. Am.* **2,** 70–76 (1954).
4. Abate, J. & Whitt, W. *Queue. Syst.* **25,** 173–233 (1997).
5. Barabási, A.-L. *Nature* **435,** 207–211 (2005).
6. Kaluza, T. *Sber. Preuss. Akad. Wiss.* **54,** 966–972 (1921).
7. Bunde, A., Eichner, J. F., Havlin, S. & Kantelhardt, J. W. *Physica A* **342,** 308–314 (2004).

Supplementary information accompanies this communication on *Nature*'s website.
Competing financial interests: declared none.
doi:10.0138/4371251a

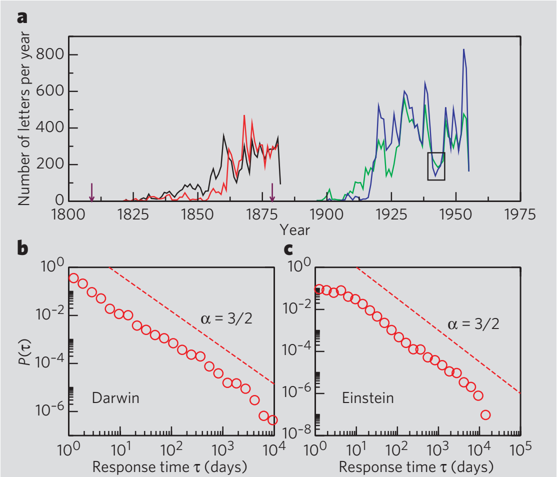

Figure 1 | The correspondence patterns of Darwin and Einstein. a, Historical record of the number of letters sent (Darwin, black; Einstein, green) and received (Darwin, red; Einstein, blue) each year by the two scientists[1,2]. An anomalous drop in Einstein's correspondence marks the Second World War period (1939–45, boxed). Arrows, birth dates of Darwin (left) and Einstein (right). **b, c,** Distribution of response times to letters by Darwin and Einstein, respectively. Note that both distributions are well approximated with a power-law tail that has an exponent $\alpha = 3/2$, the best fit over the whole data for Darwin giving $\alpha = 1.45 \pm 0.1$ and for Einstein $\alpha = 1.47 \pm 0.1$.

The classical M/M/1 process, which assumes that letters arrive at a rate λ and are answered at a rate μ, can be used to model their correspondence patterns. The waiting-time density of the M/M/1 process follows $f(t) \sim t^{-3/2} \exp(-t/t_0)$ for $\lambda \leq \mu$ [J. Abate and W. Whitt (1997), Queueing Systems 25, pp 173-233], which predicts a power-law waiting time for the critical regime $\lambda = \mu$, when $t_0 = \infty$.

However, by the response rates 0.32 and 0.24, we have $\lambda > \mu$ and this places the model in the supercritical regime, where a finite fraction of letters are never answered. Oliveira and Barabási [Nature, 437, p. 1251] pointed out that numerical simulations indicate that in this supercritical regime the waiting-time distribution of the responded letters also follows a power law with exponent $\alpha = 3/2$.

2 Why power-law?—Modeling!

Clearly, Oliveira and Barabási [Nature, 437, p. 1251] has posed such questions on modeling the correspondence patterns of human being:

1, how does the ordinary people prioritize the correspondence in need of a response? and

2, does a usual priority principle really lead to a power-law waiting time in the supercritical regime?

To partially answer the above questions, or at least, to provide some useful evidence for understanding the above questions, we introduce a special queueing system with a service discipline, which corresponds a usual priority principle of ordinary people, and then study the waiting time for served customers, especially in the supercritical regime.

3 Our model: A Non-Markovian queueing system

Now, let's consider the usual M/M/1 queue, the simplest queueing model used in practice. Suppose that the arrivals occur in a Poisson process with rate λ and the service times of the unique server have an exponential distribution with parameter μ.

We put the service discipline as follows:

1, for some fixed $T > 0$, a customer leaves the queue when his waiting time exceeds T;

2, the remains are served on the *last in first out* principle, namely, the principle serves customers one at a time, and the customer with the shortest waiting time will be served first.

Assume that at time 0, a customer (the 0-th customer) arrive and the queueing system begin to work.

Clearly, our model is a M/M/1 system with impatience, and the corresponding queueing process $\{N(t) : t \geq 0\}$ is a non-Markovian. Where $N(t)$ is the total number of customers in the system.

The research on queues with impatience was started by [Barrer (1957), OR 5], then followed by [Finch (1960)], [Gnedenko and Kovalenko (1968)], [Jurlevic (1970, 1971)], [Baccelli and Hebuterne (1981)] and [Kok and Tijms (1985)] *etc.*

For non-Markovian systems, the corresponding queue length, busy period, waiting time and loss probability are always studied under the ASSUMPTION of Statistical Equilibrium!

Here is a gap: give a proof to the "Statistical Equilibrium"!

1. Main results

Assume that at time 0, a customer (the 0-th customer) arrive and the queueing system begin to work.

For any $n \geq 0$, Denote by D_n the waiting time of the n-th customer, namely, D_n is the usual waiting time when he is finally served before his waiting time exceeds T, otherwise, $D_n = \infty$. Let

$$\tau = \tau(\lambda, \mu, T) = \inf\{n \geq 1 : D_n = 0\}.$$

Denote $q_k = \mathbb{P}(\tau = k)$, $k \geq 1$, and let $M = M(\lambda, \mu, T) = \mathbb{E}(\tau)$ be the expectation of τ.

First of all, we have the following proposition on $\mathbb{E}(\tau)$.

Proposition 1:

i) for any $\lambda,\ \mu > 0$ and $0 < T < \infty$, we have

$$M = M(\lambda, \mu, T) = \mathbb{E}(\tau) < \infty.$$

ii) for any given $\mu > 0$, $M = M(\lambda, \mu, T)$ increases in λ and T. Furthermore, for any given λ and μ,

$$\lim_{T\to\infty} M(\lambda, \mu, T) = M(\lambda, \mu, \infty) \begin{cases} < \infty, & \text{if } \lambda < \mu; \\ = \infty, & \text{if } \lambda \geq \mu. \end{cases}$$

Now, we state our main theorem as follows.

Theorem 1: For any $\lambda,\ \mu > 0$ and $0 < T \leq \infty$, W_n, the waiting time of the n-th served customer, converges in distribution to a non-negative random variable W_T. Furthermore, the distribution function F_T of W_T satisfies $F_T(x) = 0$, for $x < 0$, $F_T(x) = 1$ for $x > T$, and

$$F_T(x) = \frac{1}{C(T)}\left[\frac{1}{M} + \frac{1}{\rho \vee 1}\left(1 - \frac{1}{M}\right)\int_0^x f_\rho(t)dt\right],$$

for $0 \leq x \leq T$, where $C(T)$ is the normalization constant, $\rho = \lambda/\mu$ and

$$f_\rho(t) = \sqrt{\rho \vee \rho^{-1}}\ \frac{1}{t}\ e^{-(\lambda+\mu)t} I_1(2t\sqrt{\lambda\mu}), \quad t > 0.$$

Where $I_1(t)$ is the modified Bessel function of first kind for real and positive t given by

$$I_1(t) = \sum_{m=0}^{\infty} \frac{1}{m!(m+1)!}\left(\frac{t}{2}\right)^{2m+1} \sim \frac{e^t}{\sqrt{2\pi t}}, \text{ as t} \to \infty.$$

2. Remarks

Remark 1: The function f_ρ is the probability density of D, the length of the busy period in the classical M/M/1 system when $\rho \leq 1$, and is the conditional probability density of D conditioned on $D < \infty$ when $\rho > 1$. Note that in case of $\rho > 1$, $\mathbb{P}(D < \infty) = \rho^{-1}$.

Remark 2: By Theorem 1, one has

$$f_\rho(t) \sim t^{-\frac{3}{2}} \exp\left(-(\sqrt{\lambda} - \sqrt{\mu})^2 t\right), \quad t \to \infty.$$

So, for any large $0 < T \leq \infty$, $f_T(t)$, the density of W_T has a power-law tail with exponent $\alpha = 3/2$ in the critical regime $\lambda = \mu$ as $t \uparrow T$. In both subcritical and supercritical cases, $f_T(t)$ decays exponentially fast.

3. Possible correspondence patterns of Darwin and Einstein

But, what are the correspondence patterns of Darwin and Einstein? Numerical simulations indicate that they always kept themselves in the critical regime. A reasonable explanation may be the following: letters arrived according to a Poisson process with rate $\lambda = \lambda_1 + \lambda_2 > \mu$, where $\lambda_1 < \mu$ is the rate of letters heard from friends and family members, λ_2 is the rate of letters heard from the strangers. As the most distinguished scientists in their research fields, Darwin and Einstein received too much letters from the strangers, so they had to **ignore** such a received letter with probability $1 - (\mu - \lambda_1)/\lambda_2$ such that, in their eyes, letters arrived according to a Poisson process with rate $\lambda_1 + \lambda_2 \times \dfrac{\mu - \lambda_1}{\lambda_2} = \mu$.

1. The limit of $\mathbb{P}(D_n = 0)$

Before we give a proof to the Main Theorem, we have to prove the following lemma. Note that this lemma plays the key role in the proof of Theorem 1.

Lemma 1: Suppose that $\lambda, \mu > 0$ and $0 < T \leq \infty$, then

$$\lim_{n\to\infty} \mathbb{P}(D_n = 0) = \frac{1}{M(\lambda, \mu, T)}.$$

Proof: In case of $T = \infty$, we are dealing with a classical M/M/1 system, the lemma follows from a standard argument for birth-death process.

Now, we suppose $0 < T < \infty$.

Let $P_{00}^{(n)} = \mathbb{P}(D_n = 0)$, $n \geq 0$. Then $P_{00}^{(0)} = 1$, and for any $n \geq 1$,

$$P_{00}^{(n)} = \sum_{k=1}^{n} q_k P_{00}^{(n-k)},$$

where $q_k = \mathbb{P}(\tau = k)$. This indicates that the sequence $\{P_{00}^{(n)} : n \geq 0\}$ is iteratively determined by $\{q_k : k \geq 1\}$ and its initial value $P_{00}^{(0)} = 1$.

By the basic theory on discrete-time Markov chains, to prove the lemma, it suffices to construct a $\mathbb{Z}_+$-valued discrete-time Markov chain $\{\xi_n : n \geq 0\}$ such that $\{\xi_n\}$ is *ergodic* and

$$f_{00}^{(k)} := \mathbb{P}(\tau_0^+ = k \mid \xi_0 = 0) = q_k, \quad k \geq 1, \qquad (*)$$

where $\tau_0^+ = \inf\{n \geq 1 : \xi_n = 0\}$ is the first return time of state 0. Note that *ergodic* means *irreducible, aperiodic* and *positive recurrent* as usual.

To this end, we give the following transition matrix $P = (P_{ij})$ to $\{\xi_n\}$:

$$P_{ij} = \begin{cases} \dfrac{q_{i+1}}{1 - \sum_{k=1}^{i} q_k}, & j = 0; \\ 1 - P_{i0}, & j = i + 1; \\ 0, & \text{else}. \end{cases}$$

It is straightforward to check that $\{\xi_n\}$ with the above transition matrix P is ergodic and satisfies (*). Thus we finish the proof of the lemma. □

Proof of the main Theorem: For any $0 \le x \le T$, one has

$$\mathbb{P}(D_n \le x) = \mathbb{P}(D_n = 0) + \underline{\mathbb{P}(D_n \le x \mid D_n > 0)}\mathbb{P}(D_n > 0).$$

First of all, $D_n > 0$ if and only if, at S_n^X, the arrival time of the n-th customer, the unique server is occupied, namely, $N(S_n^X -) > 0$. It is clear that, on the *last in first out* principle, if $N(S_n^X -) > 0$, then D_n does not depend on the exact value of $N(S_n^X -)$. Second, for any $0 < x \le T$, $D_n \le x$ implies that all customers arrived after, but were served before the n-th one are finally served in time x $(\le T)$. Hence, by the *memoryless property* of the exponential distribution, for any $0 < x \le T$ and for any $n \ge 1$,

$$\underline{\mathbb{P}(D_n \le x \mid D_n > 0)} = \mathbb{P}(D \le x) =: F_D(x),$$

where D is the length of the busy period of the classical M/M/1 system.

Let $\Gamma(s)$ be the Laplace transform of $F_D(x)$ defined as the following Lebesgue-Stieltjes integration

$$\Gamma(s) = \int_0^\infty e^{-st} dF_D(t), \quad \mathrm{Re(s)} > 0,$$

where $\mathrm{Re(s)}$ is the real part of the complex number s. Then

$$\Gamma(s) = \frac{1}{2\lambda}\left[\lambda + \mu + s - \sqrt{(\lambda+\mu+s)^2 - 4\lambda\mu}\right].$$

This can be inverted to give the explicit form

$$f_D(t) = F_D'(t) = \sqrt{\frac{\mu}{\lambda}}\,\frac{1}{t}\,e^{-(\lambda+\mu)t} I_1(2t\sqrt{\lambda\mu}),$$

where $I_1(t)$ is the modified Bessel function given in the statement of the theorem.

In case of $\lambda \leq \mu$, we have $\mathbb{P}(D < \infty) = 1$, so $F_D(x)$ is a probability distribution function and $f_D(t)$ given in is a probability density.

In case of $\lambda > \mu$, let

$$\tilde{\Gamma}(s) = \frac{\lambda}{\mu}\Gamma(s) = \frac{1}{2\mu}\left[\lambda + \mu + s - \sqrt{(\lambda+\mu+s)^2 - 4\lambda\mu}\right],$$

then, by symmetry and the linearity of the Laplace Transform, its inverse $\tilde{f}_D(t) = \frac{\lambda}{\mu} f_D(t)$ is also a probability density corresponding to the following probability distribution function

$$\tilde{F}_D(x) := \frac{\lambda}{\mu} F_D(x) = \mathbb{P}(D \leq x)/\mathbb{P}(D < \infty) = \mathbb{P}(D \leq x \mid D < \infty).$$

Clearly, one has $\mathbb{P}(D < \infty) = \mu/\lambda$ and $\tilde{F}_D(x)$ is the conditional distribution function of D conditioned on $D < \infty$.

Write $f_D(t)$ and $\tilde{f}_D(t)$ in the unified form $f_\rho(t)$. Then, the conditional distribution of D_n conditioned on $D_n \leq T$ is

$F_n(x) := \mathbb{P}(D_n \leq x \mid D_n \leq T) =$

$$\frac{1}{C_n(T)} \left[\mathbb{P}(D_n = 0) + \mathbb{P}(D_n > 0) \frac{1}{\rho \vee 1} \int_0^x f_\rho(t) dt \right], \quad (**)$$

for $0 \leq x \leq T$, where $C_n(T)$ is the normalization constant. The theorem follows immediately form $(**)$ and Lemma 1. □

References:

- J. Abate and W. Whitt (1997) *Assymptotics for M/G/1 low-priority waiting time tail probabilities,* Queueing Systems 25, pp 173-233
- W. J. Anderson (1991) *Continous-Time Markov Times,* Springer-Verlag, New York
- M.-F. Chen (2004) *From Markov Chains to Non-equilibrium Particle Systems,* 2nd ed. World Scientific Publishing Company, Singapore
- M.-F. Chen and Y.-H. Mao (2007) *An Introduction to Random Processes,* (in chinese), Higher education press, ISBN 978-7-04-021778-0
- D. Gross, J. F. Shortle, J. M. Thompson and C. M. Harris (2008) *Fundamentals of Queueing Theory,* 4th ed., John Wiley

- N. L. Johnson, S. Kotz and A. W. Kemp (1993) *Univariate Discrete distributions* 2nd ed., John Wiley & Sons, Inc., ISBN 0-471-54897-9, p. 159
- T. Lindvall (1992) *Lectures on the coupling method,* John Wiley & Sons, Inc., New York
- J. G. Oliveira and A.-L. Barabási (2005) *Darwin and Einstein correspondence patterns,* Nature 437, p. 1251
- F. W. J. Olver (1972) *Bessel Functions of Interger Order,* in *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables,* edited by M. Abramowitz and I. A. Stegun, 10th printing. New York: Dover, pp 355-434
- G.-H. Xu (1980) *Random Service Systems,* (in chinese), Science Press

Thanks for your attention!