Challenges to the Assessment of Time-to-Proof of Mathematical Conjectures

Journal Article

Author(s):
Hisano, Ryohei; Sornette, Didier

Publication date:
2013

Permanent link:
https://doi.org/10.3929/ethz-b-000076397

Originally published in:
The Mathematical Intelligencer 35(4), https://doi.org/10.1007/s00283-013-9383-7

Challenges to the Assessment of Time-to-Proof of Mathematical Conjectures

RYOHEI HISANO AND DIDIER SORNETTE

The Viewpoint *column offers readers of* The Mathematical Intelligencer *the opportunity to write about any issue of interest to the international mathematical community. Disagreement and controversy are welcome. The views and opinions expressed here, however, are exclusively those of the author, and the publisher and editor-in-chief do not endorse them or accept responsibility for them.* Viewpoint *should be submitted the editor-in-chief.*

In the 27 December 2010 issue of *New Scientist*,[1] several articles discussed progress in predicting the timing of new discoveries and forecasting the future of science and technology. In particular, Samuel Arbesman and Rachel Courtland used the waiting times for solving 18 mathematical problems to estimate the probability that the "P versus NP problem" will be solved by 2024: they arrive at roughly 50% [1]. To do this, they constructed an approximate representation of the cumulative distribution of waiting times from formulation to proof (referred to as "times-to-proof" hereafter), based on this set of 18 solved mathematical problems.

Their methodology created a heated debate on the *New Scientist* website [2]. For instance, a comment posted on 3 February 2011, 10:27:11 GMT, criticized the authors' methodology, stressing that "their method of estimation looked only at problems that actually were solved," which may introduce a selection bias. A more formal attack was published in *New Scientist* on 2 February 2011 [3]. In a nutshell: using a probability distribution amounts to assuming that the underlying generating process is stationary, but stationarity may not hold over the decades and centuries corresponding to the investigated data, as the population of mathematicians has grown significantly and their theorem-proving technology has arguably improved due, e.g., to cumulative knowledge, computers, and collective work mediated by Internet and social network tools.

Nevertheless, we think the question posed by Arbesman and Courtland is interesting. Not only is it an attempt to guess when an unsolved problem such as the P versus NP conjecture might be settled, but it also raises the issue of the evolution of productivity of mathematics throughout history. In this spirit, we revisit this question and analyze a larger database of 144 conjectures including both closed and open conjectures.

But first, we assure you that we are well aware of the main caveats with attempting a statistical quantification of the generation of mathematical results during its (relatively recent) history.

First, any assessment of the time-to-proof distribution of mathematical conjectures can be criticized as being meaningless if it ignores their content and context as well as several other issues. Specifically,

1) A "time-to-proof" depends on the content of the conjecture. Related to the content is the question of whether mathematicians judge the conjecture worth pursuing, and why. This is not unique to mathematics. Indeed, consider a typical individual (a mathematician in our context) who is subjected to a flow of information

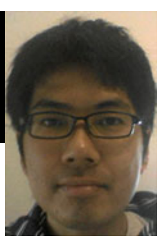
[1] http://www.newscientist.com/article/mg20827923.700-2011-preview-milliondollar-mathematics-problem.html.

DOI 10.1007/s00283-013-9383-7

(research papers to read, conferences she is attending, visits of colleagues, and so on) and requested tasks (teaching, administration, etc.), under time, energy, regulatory, social, and monetary constraints. She will respond by a sequence of actions that themselves contribute to the flow of influences spreading to other mathematicians. Remarkably, stationary distributions have been documented quantitatively for the waiting times between triggering factor and response of a number of human activities, such as the waiting times until an e-mail message is answered [4], the time intervals between consecutive e-mails sent by a single user and time delays for e-mail replies [5], the waiting time between receipt and response in the correspondence of Darwin and of Einstein [6], and the waiting times associated with web browsing, library visits, and stock trading [7]. In each of these activities, one could forcefully argue that the reported distributions may be meaningless, aggregating "carrots" and "potatoes," because each single different human activity is strongly influenced by its specific content and the proximate interest it represents to its user. Yet, the evidence suggests a kind of universal behavior that is worth investigating, even for mathematical conjectures. Despite the variability in the characteristics of conjectures (and of other human problems and activities), there might be a homogenous process underlying the generation of the problems and their resolutions. In this spirit, although the underlying generating mechanism(s) of the purported distributions are not known for certain, we suggest that recent modeling progress based on priority queuing theory may be relevant [8, 9]. Another theoretical approach consists in thinking of mathematical research as a bundle of random walks in some high-dimensional mathematical space, such that intersections between them or crossing of some boundary corresponds to a successful outcome and the establishment of the proof. Such models have been argued to apply for instance to the space of investment strategies used by a large populations of traders, explaining the long memory of financial volatility as resulting from the statistical properties of random walk crossing in arbitrary spaces [10].

2) In addition, the definition of what constitutes a "proof" has changed with time, and, within a given definition paradigm, proofs hold different standings. For instance,

AUTHORS



RYOHEI HISANO obtained his M.A. in Economics from Hitotsubashi University and is currently a Ph.D. candidate in the Department of Management, Technology and Economics at ETH Zurich. His work focuses on analyzing large-scale socio-economic data using interdisciplinary approaches at the crossing between complexity science, physics, economics, and computational intelligence. Besides work, he enjoys traveling and taking photographs of the places he has visited.

D-MTEC
ETH Zurich
Scheuchzerstrasse 7
CH-8092 Zurich
Switzerland
e-mail: em072010@yahoo.co.jp

DIDIER SORNETTE is professor of Entrepreneurial Risks in the Department of Management, Technology and Economics at ETH Zurich and is a professor of finance at the Swiss Finance Institute. He is also an associate member of the Department of Physics and of the Department of Earth Sciences at ETH. He is the author many research papers and seven books. Much of Professor Sornette's research focuses on the prediction of crises and extreme events in complex systems, and in particular of financial bubbles and crashes, and the diagnostic of systemic instabilities. Other applications include earthquake physics and geophysics, financial economics and the theory of complex systems, the dynamics of success in social networks, and the complex system approach to medicine (immune system, epilepsy…). The many subjects and disciplines he is working on are like hobbies. When "tired" of one subject, he finds refreshment and rejuvenation in shifting to another scientific challenge.

D-MTEC
ETH Zurich
Scheuchzerstrasse 7
CH-8092 Zurich
Switzerland
e-mail: dsornette@ethz.ch

consider the famous first proof [11, 12] of the four-color theorem. No human or group of humans has currently verified this proof in full, and hence some may not consider it fully proven. Its status has in fact been evolving through time, with the initial lengthy computer-assisted proof [11, 12] being confirmed later by simpler proofs that rely on well-tested general purpose theorem-proving software [13]. This raises the question of which proof to use to measure a time-to-proof. There are also a variety of proof types, such as constructive and nonconstructive. We propose to refer to the "wisdom of crowds" or, more precisely, the wisdom of the mathematical community, which has built up collectively the list of conjectures and their proofs, reported in the free web encyclopedia Wikipedia. We comment later on the quality of this dataset.

3) Finally, some famous conjectures might in fact be undecidable. This corresponds to the situation where there is a nonzero probability mass (atom) at infinity in the distribution of time-to-proof, a situation that is generic in priority queuing models when the average arrival rate of tasks to perform is larger than the average rate of performing or solving these tasks [8, 9]. This has also been documented empirically in the distribution of waiting times exhibited by human users before upgrading computer programs [14]. Thus, our expression (7) below has to be restricted to the set of decidable open conjectures, which we do not know a priori. Thus, deviations from relation (7) by the empirical distributions of closed and open conjectures may be taken as diagnostics of the existence of such undecidable conjectures. In the following, we discuss other issues involved with (7) resulting from the additional and arguably dominating nonstationarity.

With these caveats, we use the larger Wikipedia database of 144 conjectures including both closed and open conjectures to dissect the major problems associated with the inference of the distribution of times-to-proof for mathematical conjectures, given the available data and the intrinsic nonstationarity of the system. We show that, even under the naïve assumption of constant average productivity per mathematician, because of the approximate exponential growth of the mathematician population, the true time-to-proof distribution is hidden, preventing us from analyzing it directly. Moreover, even under the assumption of a constant average productivity per mathematician, we could not reject the simplest model of an exponential rate of conjecture proof with a rate of 0.01/year for the dataset (translating into an average waiting time to proof of 100 years). Our analysis highlights the major challenges behind applying quantitative methods to the assessment of time-to-proof.

First, we present the dataset. Then, we review and adapt the theory of recurrence processes to the distribution of time-to-proof. In subsequent sections, we present the empirical distribution of time-to-proof obtained from our dataset, formulate the consequences of the nonstationarity of the births of conjectures, and combine this nonstationarity with different models for the intrinsic distribution of time-to-proof to fit the empirical distributions. Finally, we stress the caveats of the proposed analysis.

Dataset

The dataset that the mathematical community has collectively contributed in constructing the page "list of conjectures" in Wikipedia, the free web encyclopedia, consists of about 160 proved and unsolved conjectures [15]. Although it is difficult, if not impossible, to establish that this list is representative and does not represent a biased sample, other tests have shown that the accuracy of Wikipedia's articles compares well with that of the standard, Encyclopaedia Britannica [16]. We think it better to try to work with what is available than to do nothing. At least we may learn something about the limitations that we need to overcome. And even a small partial insight toward this grail of characterizing the process of mathematical creativity and production is worth trying. Perhaps this article will encourage the community to develop a more extensive database of mathematical conjectures, in the spirit of the Erdös Number Project [17] in Graph Theory, which studies research collaboration among mathematicians.

Distribution of Time-to-Proof: Definition and Theory

For each conjecture i present in the "list of conjectures" in Wikipedia, we searched for the exact year t_1^i when it was stated and the exact year t_2^i when it was resolved (or whether it still remains open), always striving to obtain the first original source or reference. For 16 conjectures, we were unable to determine the exact values of t_1^i and/or t_2^i, thus reducing our usable dataset to 144 conjectures, of which 60 have been solved so far (January 2012) and 84 are still open problems.

We determine the time-to-proof τ_i^c for each of the 60 conjectures that have been solved (or proven wrong) using the formula

$$\tau_i^c := t_2^i - t_1^i \quad (1)$$

For the 84 open conjectures, the relevant variables are the so-called "backward recurrence times" defined by

$$\tau_i^b := t - t_1^i \quad (2)$$

where t is 2012. We study the complementary cumulative distribution functions (ccdf) (also called "survivor functions") $S_c(\tau)$ and $S_b(\tau)$ corresponding to the times-to-proof τ_i^c and τ_i^b, respectively.

We will assume that the generating mechanisms for the waiting times to proof of closed and open conjectures are similar. At first glance, one might think that closed and open conjectures form two distinct classes, with closed conjectures easier to solve because they have been solved. This argument may be valid for conjectures that have been closed very fast, say on time scales ranging from months to a few years. However, for conjectures that we examine that have taken decades or centuries to close, we find this argument less compelling. Indeed, any conjecture that takes decades or more to close forms a kind of bold extension to current knowledge and to existing mathematical "technology." The formulation of such a challenging new conjecture is an intrinsic part of the research and creative process. We propose that the reason open conjectures still remain open is because tools and ideas have not yet reached the threshold needed for their solution.

Moreover, we assume more boldly that the average productivity per mathematician contributing to this discovery procedure is constant throughout the years. This may or may not be the actual case. However, with this assumption we can turn to the other major challenges (beside data availability) that underlie the application of any quantitative methods to the assessment of the time-to-proof distribution.

The mathematical theory of interval distributions for stationary point processes provides an exact correspondence between the survival functions of closed conjectures and open conjectures that we can use when interpreting the empirical distributions. Defining $N(\tau, \tau+t]$ as the number of events in the time interval $(\tau, \tau+t]$ (which excludes the left side and includes the right side of the time interval), the backward recurrence time of an event generated by a point process, is defined formally as

$$\tau_t := \inf\{u > 0 : N(t-u, t] > 0\} \qquad (3)$$

where $t = 2012$ [18]. In words, it is the time interval from the latest event (the formulation of a conjecture) to present (at which time the conjecture is still open), such that there is one event in this interval. For a stationary process, we have the identity [19]

$$\begin{aligned} &Pr\{N(0,\tau) \geq 1, N(\tau, \tau+t) = 0\} \\ &= Pr\{N(\tau, \tau+t) = 0\} - Pr\{N(0, \tau+t) = 0\} \end{aligned} \qquad (4)$$

In words, the probability $Pr\{N(\tau, \tau+t) = 0\}$ that there are no events in $(\tau, \tau+t)$ is equal to the probability $Pr\{N(0, \tau+t) = 0\}$ that there are no events in $(0, \tau+t)$ plus the probability $Pr\{N(0,\tau) \geq 1, N(\tau, \tau+t) = 0\}$ that there are no events in $(\tau, \tau+t)$ and at the same time there is at least one event in $(0, \tau)$. In other words, the fact that the interval $(\tau, \tau+t)$ has no event can be associated with the occurrence of either no event or of some events earlier in $(0, \tau)$. Dividing both sides by τ, taking the limit $\tau \Rightarrow 0$ of expression (4), and using the definitions

$$S_c(t) = lim_{\tau \longrightarrow 0} Pr\{N(\tau, \tau+t) = 0 | N(0,\tau) \geq 1\} \qquad (5)$$

and

$$S_b(t) = Pr\{N(0,t) = 0\} \qquad (6)$$

for the complementary cumulative distribution functions $S_c(\tau)$ and $S_b(\tau)$ corresponding to the times-to-proof τ_i^c and τ_i^b, respectively defined by (1) and (2), identity (5) translates into the Palm-Khinchin relation [18–20]

$$S_b(t) = -\frac{1}{\lambda}\frac{d}{dt} S_c(t) \qquad (7)$$

where λ is the inverse of the average time-to-proof (i.e., $\lambda = lim_{\tau \longrightarrow 0} Pr\{N(0,\tau) = 1\}/\tau$). The conditioning in (5) ensures that the counting of the time to the next event is indeed starting from the previous one (the condition $N(0,\tau) \geq 1$).

Empirical Distributions of Time-to-Proof

The birth flow of mathematical conjectures is not uniformly distributed with time. For the dataset of 60 solved conjectures, Figure 1 shows a scatter plot, with the abscissa giving the year when the conjecture was stated and the ordinate the year when it was solved. The time axes cover the period from 1600 CE to present. We can see that the plot is significantly crowded at the upper right part of the figure. This implies that the birth flow of conjectures is increasing with time.

Figure 2 quantifies this visual impression by showing the cumulative number of stated problems $N_{closed}(t)$ (that

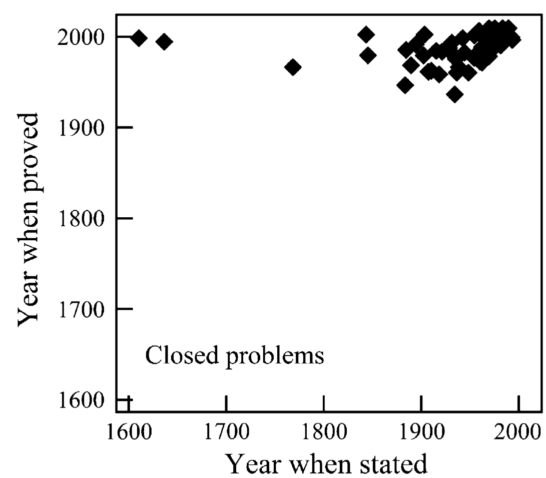


Figure 1. Scatter plot showing when a mathematical problem was stated and when it was resolved for the 59 closed problems (excluding the honeycomb conjecture) in our dataset, starting from the year 1600.

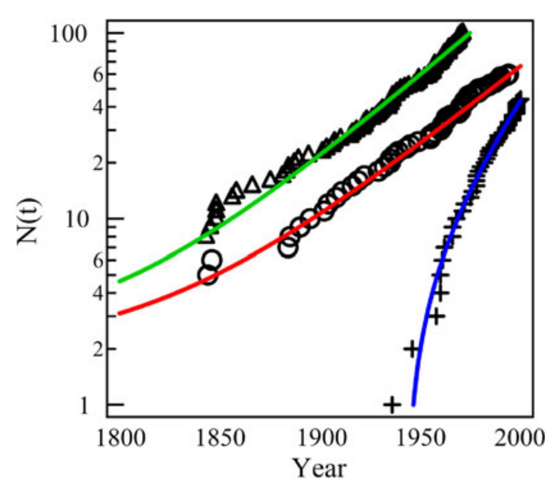


Figure 2. Cumulative number of stated problems $N_{closed}(t)$ that have found a solution (*circles*), cumulative number of stated problems $N_{stated}(t)$ that are both closed and still open (*triangles*), and cumulative number $N_{solution}(t)$ of the solutions of the solved problems (*crosses*) from 1850 CE to 2000. The continuous lines correspond to the exponential growth models (8–10).

Table 1. Summary of the estimated parameters for equations (8) to (10)

	a	b	c
N_closed	0.02 [0.019,0.021]	1790 [1778,1802]	1.9 [0.78,3.02]
N_stated	0.02 [0.019,0.021]	1875 [1865,1885]	2.3 [−0.09,4.7]
N_solution	0.035 [0.031,0.039]	1902 [1889,1915]	−6.5 [−9.4,−3.57]

The fitted equation is $N(t) = \exp[a * (t - b)] + c$. Numbers in square brackets represent 95% confidence intervals.

have found a solution) and $N_{stated}(t)$ (that are both closed and still open) and the cumulative number $N_{solution}(t)$ of the solved problems from 1850 CE to 2000. Exponential growth models fit rather well the different data sets. The best fits to these three data sets give, respectively

$$N_{closed}(t) = \exp[0.02 * (t - 1790)] + 1.9 \quad (8)$$

$$N_{stated}(t) = \exp[0.02 * (t - 1875)] + 2.3 \quad (9)$$

$$N_{solution}(t) = \exp[0.035 * (t - 1902)] - 6.5 \quad (10)$$

where t is given in units of years and is counted since the beginning of the present era. Table 1 summarizes the estimated parameters together with their 95% confidence intervals.

The average growth rate of the number of new conjectures is approximately equal to 0.02 year^{-1}, corresponding to a tripling of the number of new conjectures every 55 years. This growth rate is close to the average growth rate of the world human population over the same period, as shown in Figure 3. The best exponential fit to the world population from 1750 to present (data retrieved from the United Nations website [21]) is

$$N(t) = \exp[0.018 * (t - 1905.6)] + 0.552 \quad (11)$$

Taking the growth of the world population as a proxy for the growth of the number of mathematicians (though this may underestimate the true number of mathematicians), we see that the average growth rate of the number of new conjectures is closely tied to the increase of the population of mathematicians. This average exponential law (11) is only a first-order approximation, as it is well known that the growth rate of the world population has varied significantly during the last few centuries [22–24]. However, given the coarse-grained nature of our dataset on mathematical conjectures, the average exponential growth (11) provides a reasonable first representation of the nonstationarity resulting from the increase of the population of mathematicians.

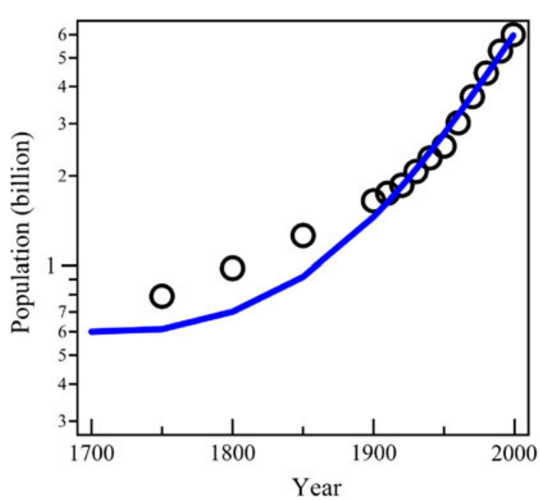


Figure 3. Growth of the world population from 1750 to present (data retrieved from the United Nations website [6]), taken as the simplest proxy for the growth of the relevant population of mathematicians. This growth can be reasonably approximated by an exponential growth given by expression (11) (*continuous line*).

The Consequence of Increasing Birth Flow

The exponential growth of the birth flow of conjectures implies that the observable distribution of times-to-proof is bounded by an exponential distribution with rate 0.02. In other words, it cannot decay more slowly asymptotically than an exponential with rate 0.02. In terms of a CCDF plot, this implies that the empirical times-to-proof distribution of closed and open problems lies to the left of this exponential distribution. To see this, write the distribution $P(\tau)$ of waiting times between formulation and proof in terms of the rate $r(t_1) = r_0 e^{at_1}$ of conjecture formulations and of the conditional distribution $p(t_2|t_1) = f(t_2 - t_1)$ that the conjecture will be proved at t_2 given that it has been formulated at t_1. We assume a constant growth rate a for $r(t_1)$ and stationarity for $p(t_2|t_1)$. This second condition provides an upper bound for the distribution. In other words, the true distribution will decay at least as fast as derived from the assumption of stationarity of $p(t_2|t_1)$. We have

$$P(\tau) = \int_0^t dt_1 \int_0^t dt_2 r(t_1) p(t_2|t_1) \delta(t_2 - t_1 - \tau) = r_0 \int_0^t dt_2 e^{a(t_2-\tau)} f(\tau) = Ce^{-a\tau} f(\tau) \quad (12)$$

where

$$C = r_0 \int_0^t dt_2 e^{at_2} \quad (13)$$

Thus, $P(\tau)$ decays no more slowly than e^{-a^τ}, that is, proportionally to the inverse of the rate of conjecture births.

Figure 4 plots the CCDF of the times-to-proof of closed and open conjectures defined by equations (1) and (2), together with the exponential bound derived above. Comparing the three distributions, we see that the CCDF for the open conjectures lies above the exponential bound. This implies that a significant number of conjectures are likely to be still missing in the bulk of the distribution. In other words, there are many missing conjectures with intermediate values of their time-to-proof, compared with the conjectures with extremely large waiting times. Figure 5 depicts the empirical distribution obtained by removing all times-to-proof smaller than 20 years, that is, by introducing

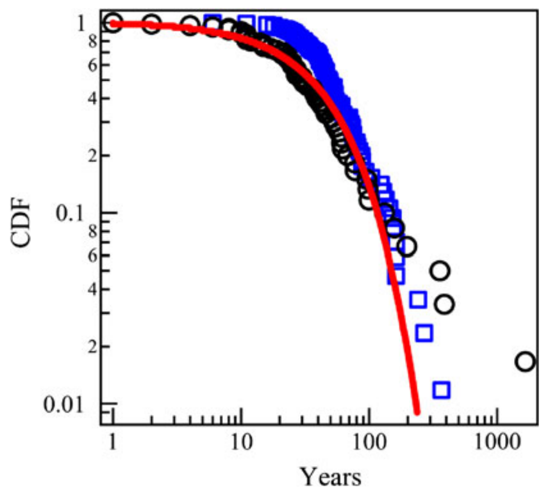


Figure 4. Complementary cumulative distribution functions (ccdf) of the predicted times-to-proof for open problems (*rectangles*), closed problems (*circles*), and an exponential distribution with rate 0.02 (*continuous line*).

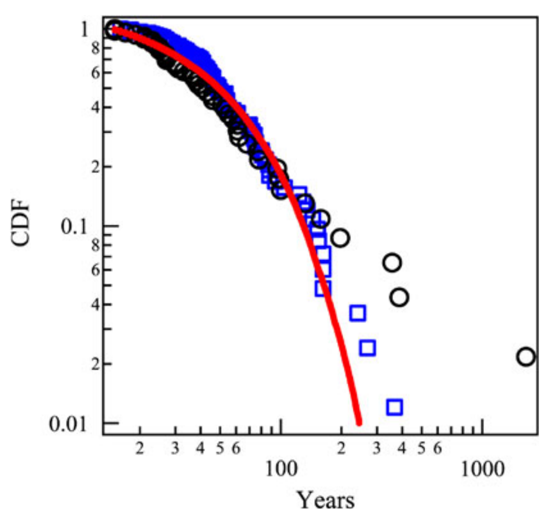


Figure 5. Complementary cumulative distribution functions (ccdf) of the times-to-proof for open problems (*rectangles*), closed problems (*circles*), and an exponential distribution with rate 0.02 (*continuous line*), obtained by introducing a lower threshold equal to 20 years, that is, by removing all times-to-proof smaller than 20 years.

a lower threshold, so that we can hope that the number of missing conjectures is reduced and the data are less incomplete. We can see that the time-to-proof of closed and open conjectures lies just on the exponential bound, except for the four largest data points to which we will return later. Pushing the threshold above 20 years does not change this behavior that both CCDF's lie on the exponential bound. The fact that the two distributions coincide confirms that the underlying distribution of time-to-proof is asymptotically close to an exponential distribution.

Simulation Analysis

We now attempt to find distributions of time-to-proof and their associated parameters that could generate the distributions shown in Figure 5. We first generate a set of N instants t_1^i, $i = 1, \ldots, N$, corresponding to the formulation times of N mathematical problems. These N times are sampled according to a Poisson process with an intensity growing exponentially with the rate 0.02 year^{-1}, obtained from the fits shown in Figure 2. This generation of conjectures mimics the structure of our dataset. This reflects a scenario in which each mathematician generates on average the same number of conjectures per unit time, while the number of mathematicians increases roughly exponentially in parallel with the growth of the human population. A naïve approach would go as follows. For each conjecture inception time t_1^i, we draw a random number τ_i^c corresponding to the time-to-proof of this conjecture. This random number τ_i^c is generated by using an intrinsic distribution associated with the way mathematics would be practiced by a population of mathematicians of constant size, technology, and mental prowess.

We have constructed synthetic catalogues of conjectures with their birth and proof times, using four different families of distributions, namely exponential, lognormal, inverse Gaussian and Burr type-III distribution. For each of these four families of distributions, the parameters were set so as to fit the empirical distribution as closely as possible and, at the same time, to reproduce the ratio of the number of closed to open conjectures (i.e., 42:80). For the exponential family, we find that the rate $\lambda = 0.01$/year provides the best fit. For the lognormal distribution, the best parameters correspond to a log-average of $\mu = 4.2$, and standard deviation $\sigma = 1.2$. For the inverse Gaussian distribution,

$$f_{ig}(x) = \left[\frac{\lambda}{2\pi x^3}\right]^{0.5} \exp\frac{-\lambda(x-\mu)^2}{2\mu^2 x} \tag{14}$$

the best parameters are $\lambda = 170$, $\mu = 72.25$. For the Burr distribution

$$pdf_{BurrIII}(x) = \frac{cd}{x^{c+1}(1+x^{-c})^{d+1}} \tag{15}$$

the best parameters are $c = 0.5$, $d = 3$. As was expected from the result of the previous section, the goodness of fit of the distributions seems to be slightly better for distributions that asymptotically behave like an exponential distribution rather than a power-law distribution. For this reason, we show only the distributions for the exponential family together with the empirical distributions in Figure 6.

The paucity of our dataset makes it nearly impossible to distinguish whether one of these distributions provides a better fit than any other distribution. Using the exponential distribution corresponds to following Occam's razor of parsimony with the simplest model providing the best fit. This suggests that most of the conjectures in our database can be described approximately by an underlying waiting time-to-proof distribution that is an exponential distribution with rate of 0.01/year.

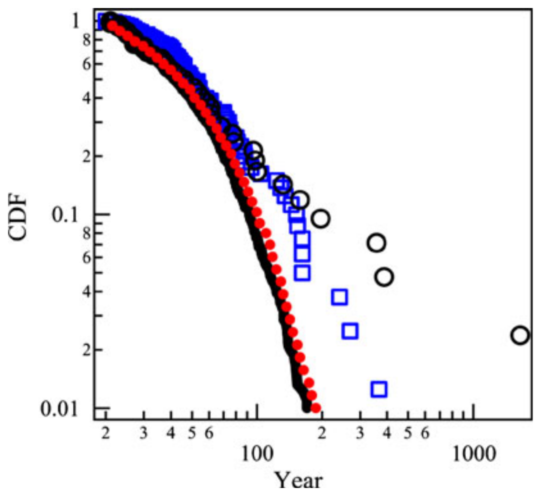


Figure 6. Comparison between the empirical cumulative distribution functions (ccdf) of the times-to-proof for open problems (*open squares*), closed problems (*open circles*), and the ccdfs generated by the simple model presented in the text using an exponential distribution of the intrinsic time-to-proofs with rate $\lambda = 0.01$/year (black curve: closed problems; full red circles: open problems).

Concluding Remarks

As we dug into this statistical analysis, we realized the need to take into account the strong nonstationarity of the problem.

As Figure 6 suggests, our best model is not the whole story. In particular, a half-dozen closed conjectures depart rather significantly from the proposed best model. One possible reason for this deviation is that the assumption of an exponential growth of the rate of conjecture births may be too simple due to the known deviation of the human population growth from a simple exponential process [22–24]. A more realistic birth-flow model of conjectures would be to observe the impact of a possible increase in mathematical productivity.

Another question is the incompleteness of the available dataset, in particular of the likely severe undersampling of the many conjectures whose time-to-proof is in the range of years to a few decades. Only conjectures that have resisted mathematicians' assaults or have played particularly distinguished and meaningful roles in the structure and history of mathematics are likely to acquire the status and fame to be recorded in databases such as the one we have used. We also neglected the possibility that the average productivity per mathematician might not be stationary. With the advances of modern technology, assuming that nonstationarity stems only from the increase in the population of mathematicians might be overly simplistic. However, with our limited data, it is impossible to disentangle this factor from the growth of the mathematician population. These remarks illustrate the difficulties associated with any attempt to extract the distribution of time-to-proof that would really show the intrinsic productivity of mathematicians throughout history.

To conclude, notwithstanding all these difficulties and caveats, if we have to make a best guess and revisit the question first raised by Arbesman and Courtland (2010), we can use the exponential distribution with rate 0.01/year together with the exponential growth of the mathematician population to calculate the probability that the "P versus NP problem" will be solved by the year 2024. We calculate the value 41.3% (with the 95% confidence interval being [38.3; 44.4%]). This suggests that Arbesman's and Courtland's original estimate of a 50% chance [1] was somewhat optimistic but was still of the right order of magnitude.

REFERENCES

[1] Arbesman, S. and Courtland, R. (25 December 2010) 2011 preview: million-dollar mathematics problem, *New Sci.* 2792.

[2] http://www.newscientist.com/article/mg20827923.700-2011-preview-milliondollar-mathematics-problem.html.

[3] Bolton, D. (2 Feb 2011) Time to P = NP, *New Sci.* 2798.

[4] Eckmann, J.-P., Moses, E., and Sergi, D. (2004) Entropy of dialogues creates coherent structures in e-mail traffic, *Proc. Natl. Acad. Sci. U.S.A.* 101, 14333.

[5] Barabasi, A.-L. (2005) The origin of bursts and heavy tails in human dynamics, *Nature* 435, 207.

[6] Oliveira, J. G. and Barabasi, A.-L. (2005) Darwin and Einstein correspondence patterns, *Nature* 437, 1251.

[7] Vazquez, A., Gama Oliveira, J., Dezso, Z., Goh, K. I., Kondor, I., and Barabasi, A. L. (2006) Modeling bursts and heavy tails in human dynamics, *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 73, 036127.

[8] Grinstein, G. and Linsker, R. (2008) Power-law and exponential tails in a stochastic priority-based model queue, *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 77, 012101.

[9] Saichev, A. and Sornette, D. (2009) Effects of diversity and procrastination in priority queuing theory: the different power law regimes, *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 81, 016108.

[10] Bouchaud, J-P., Giardina, I., and Mzard, M. (2001) On a universal mechanism for long-range volatility correlations, *Quant. Financ.* 1, 212–216.

[11] Appel, K. and Haken, W. (1977) Every planar map is four colorable, part I. discharging, *Illinois J. Math.* 21, 429–490.

[12] Appel, K., Haken, W., and Koch, J. (1977) Every planar map is four colorable, part II. reducibility, *Illinois J. Math.* 21, 491–567.

[13] Appel, K. and Haken, W. (1989) *Every Planar Map is Four-Colorable*, Providence, RI: American Mathematical Society.

[14] Maillart, M., Sornette, D., Frei, S., Duebendorfer, T., and Saichev, A. (2011) Quantification of deviations from rationality from heavy-tails in human dynamics, *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 83, 056101.

[15] http://en.wikipedia.org/wiki/List_of_conjectures.

[16] Giles, J. (2005) Internet encyclopedias go head to head, *Nature* 438, 900–901.

[17] http://www.oakland.edu/enp/ (maintained by Jerry Grossman at Oakland University).

[18] Daly, D. J. and Vere-Jones, D. (2003) *An Introduction to the Theory of Point Processes*, Vol. 1, 2nd ed., New York, Berlin, Heidelberg: Springer.

[19] Laurance, A. J. (1974) Theory of interevent interval distributions for stationary point processes, *Inform. Control* 25, 299–316.

[20] Vere-Jones, D. (1970) Stochastic Models for Earthquake Occurrence, *J. R. Stat. Soc. B* 32, 1–62.

[21] http://www.un.org/esa/population/publications/sixbillion/sixbilpart1.pdf.

[22] Johansen, A. and Sornette, D. (2001) Finite-time Singularity in the Dynamics of the World Population and Economic Indices. *Physica A* 294, 465–502.

[23] Bloom, D. E. (2011) 7 Billion and counting, *Science* 333, 562–569.

[24] Huesler, A. D. and Sornette, D. (2011) *Evidence for super-exponentially accelerating atmospheric carbon dioxide growth*, submitted (http://arxiv.org/abs/1101.2832).