Superstars without talent?
The Yule distribution controversy

Laura Spierdijk[a] and Mark Voorneveld[b, c,1]

[a] *Dept. of Econometrics, University of Groningen, The Netherlands*
[b] *Dept. of Econometrics and Operations Research, Tilburg University, The Netherlands*
[c] *Dept. of Economics, Stockholm School of Economics, Sweden*

Abstract. Chung and Cox (1994) provided an intuitively appealing stochastic model which indicates that superstars may exist regardless of talent and which gives rise to the Yule distribution. We adopt a different empirical approach and test its goodness-of-fit using a parametric bootstrap and several powerful test statistics. Just like the discrete Pareto distribution, it is overwhelmingly rejected: it is a fairly accurate approximation of the lower quantiles of the superstar distribution, but overestimates the snowball effect that makes consumers purchase records of the most successful artists. In other words, the Yule distribution captures stardom, but not superstardom. A generalization of the Yule distribution provides an excellent fit in two of the three data sets.

JEL classifications: J31, L82, Z11

---

1. Introduction

The superstar phenomenon refers to a situation where relatively small numbers of people earn enormous amounts of money and dominate the field in which they are active (art, sport, media, ...). Adler (2006) provides an overview of the economic literature on superstars, which contains a natural dichotomy. Some authors, following Rosen (1981), ascribe the phenomenon to differences in talent. Others, starting from Adler (1985), indicate that superstars may exist regardless of talent, simply because enjoying, say, music has an important social aspect and people tend to follow the crowd, thus creating a snowball effect.

Chung and Cox (1994) conduct a test of the latter variety. They adopt an intuitively appealing stochastic process by Nobel laureate Herbert Simon (1955), who in his turn was inspired by Yule (1924). Their data concerns the number of Gold Records earned by performers in the American popular music industry and the process is based on two assumptions about the purchasing behavior of consumers. Roughly speaking (see Simon, 1955, pp. 426–433 for details), consumers make consecutive purchases, such that:

> [**A1**] The probability that consumer $n + 1$ buys a record already chosen by exactly k of the previous n consumers is proportional to k.

> [**A2**] There is a constant probability $\delta \in (0, 1)$ that consumer $n + 1$ buys a record that was not previously chosen.

Assumption [A1] models the snowball effect, whereas [A2] makes the process nontrivial: it causes the support of the probability distribution to consist of the positive integers. For later reference ([C1] below), notice that this is not the case if $\delta = 0$: everybody would buy the same record as the first consumer, creating a single, infinitely successful artist. Under stationarity, Simon (1955) shows that [A1] and [A2] lead to the Yule distribution, a one-parameter (say ρ) distribution with probability mass function

$$p_k = \rho B(k, \rho + 1) \qquad (k = 1, 2, \ldots), \tag{1}$$

where $B(\cdot, \cdot)$ denotes the standard Beta function and ρ is related to δ in [A2] via

$$\rho = 1/(1 - \delta). \tag{2}$$

There are contradictory conclusions as to whether the Yule distribution provides a good description of stardom in the American popular music industry. Chung and Cox (1994) claim that it provides "an excellent description", but base this statement on the fit of an

approximation, not the Yule distribution itself; see [C3] below. Giles' (2006) analysis of recordings reaching the Number 1 position on the Billboard Hot 100 chart follows their approach and leads to the opposite conclusion, i.e., rejection of the Yule distribution as a suitable description of the data. So the Yule controversy remains: does this distribution provide a good description of superstar data or not?

Our way of testing the goodness-of-fit of the Yule distribution differs significantly from Chung and Cox (1994). We provide three short comments on their analysis. They adopt the specific parameter value $\rho = 1$, because "Simon (1955) suggests that the Yule distribution provides a good fit to various empirical data particularly when the value of ρ is equal to one". This is part of our first comment:

> [C1] The authors do not test whether the Yule distribution in general fits the data, only whether a single example of it does. But the suggested parameter value $\rho = 1$ is implausible, both for economic and mathematical reasons.

Indeed, using (2), $\rho = 1$ implies that $\delta = 0$, so that assumption [A2] required in Simon's derivation of the Yule distribution is violated:[2] nobody would ever buy a previously unchosen record. Moreover, the Yule distribution with $\rho = 1$ has infinite expectation, which is unrealistic for a measure of success/stardom. We therefore assess the goodness-of-fit for two intuitive parameter choices: the maximum likelihood (ML) estimator and a method of moment (MM) estimator.

Even if one were to accept the value $\rho = 1$ as a reasonable candidate, the two tests of Chung and Cox (1994) to assess the goodness-of-fit are subject to critique.

> [C2] The first test, a Chi-square goodness-of-fit test, is not a suitable test for the superstar phenomenon.

---

[2]In an e-mail response to our comment, professor Chung writes: "If someone raises an issue with our study based on its assumption of $\rho = 1$, we would just cite Friedman (1953):

> the relevant question to ask about the 'assumptions' of a theory is not whether they are descriptively 'realistic,' for they never are, but whether they are sufficiently good approximations for the purpose in hand. And this question can be answered only by seeing whether the theory works, which means whether it yields sufficiently accurate predictions.

In our view, $\rho = 1$ was sufficiently a good approximation for the purpose in hand because it yielded good predictions, i.e., our equation (3)."

Friedman's quote does not address the inconsistency of the authors' assumptions. Moreover, we stress that also the authors' tests of the fit of the Yule distribution with $\rho = 1$ are subject to critique; see [C2] and [C3]. Our Section 2 shows that the distribution with $\rho = 1$ and even with more intuitive parameters (the maximum likelihood estimator and a moment estimator) is overwhelmingly rejected as a suitable description of the data.

Indeed, the Chi-square goodness-of-fit test is an 'omnibus test' with little power (Moore, 1986) and the asymptotic distribution of the test statistic is only Chi-square when the expected number of observations in each group is at least five. But the predicted number of performers with a large number of Gold Records, the real superstars, is low. Such extreme observations must then be grouped together or ignored. Chung and Cox (1994, p. 774, Column 1, lines 14–19) do the latter. However, neither option is very elegant, since it essentially ignores exactly that part of the data set where the superstars are located.

Therefore, it would be better to consider statistical tests that (1) have more power and (2) attach more value to the tail of the distribution where the genuine superstars are located. Since the asymptotic distribution of goodness-of-fit tests is often only known when based on continuous data with fixed (rather than estimated) parameters, a parametric bootstrap can be used to obtain critical values in the current situation where we have a discrete distribution with estimated parameters.

[**C3**] The second test is not a test of the fit of the Yule distribution, but of an approximation, namely a power law.

Indeed, Chung and Cox (1994, formula (9)) use that $\Gamma(k)/\Gamma(k+c) \approx 1/k^c$ as long as k is much greater than c to approximate the Yule distribution by a power law. Taking $c = \rho+1$, k must be much greater than 2 for this approximation to be reliable when $\rho = 1$. However, two-thirds of their data have $k = 1$ or $k = 2$.[3] The fit of the power law is then tested via linear regression.[4] Modern computers are sufficiently fast for computations with the Gamma function, making approximations unnecessary: our statistical analysis is conducted for the Yule distribution, not for some approximation.

The stochastic model proposed by Chung and Cox (1994) is intuitively appealing, but comments [C1] to [C3] should be kept in mind when reading the literature: this initial article is often cited[5] and set the stage for more recent studies following their line of reasoning and therefore subject to the same comments. These include Chung and Cox (1998) on the movie industry, Cox and Kleiman (2000) on the finance sector, and Giles (2006), again on the popular music industry.

---

[3]In the Weeks and Hits data of Giles (2006) summarized in our Table 1, this holds for 61 and 71 percent of the data, respectively.

[4]In an e-mail response to our comment, professor Chung writes: "Please note that we also obtained the empirical estimate of rho and tested whether it is different from one. We found that rho is not statistically different from one (see footnote 7)." But their test is not related to the Yule distribution, only to an approximation.

[5]The section on empirical tests of superstardom in a recent overview of Adler (2006) gives Chung and Cox (1994) a prominent place and an Internet search on their article's main title gives dozens of hits.

We briefly relate the Yule distribution to the more widely known Pareto distribution: the discrete Pareto distribution with parameter $\rho > 0$ has probability mass function

$$q_k = k^{-\rho}/\zeta(\rho) \qquad (k = 1, 2, \ldots), \tag{3}$$

where $\zeta(\cdot)$ is the zeta function. For large k, writing $\Gamma(k)/\Gamma(k + c) \approx 1/k^c$ in (1) gives

$$p_k = \rho B(k, \rho + 1) \propto \Gamma(k)/\Gamma(k + \rho + 1) \approx k^{-(\rho+1)}, \tag{4}$$

i.e., the tail of the Yule distribution resembles a Pareto distribution. Despite similar tails, the Yule and Pareto distributions are distinct and it is impossible to express probabilities of one of them as functions or limiting values of the other. Consul (1991) gives a more detailed comparison. Log-log rank size plots of the data are not remotely linear and the hypothesis that a Pareto distribution fits the data is easily rejected by formal tests.

There is a two-parameter variant of the Yule distribution (henceforth the generalized Yule distribution) with probability mass function

$$r_k = \rho(1 - \alpha^\rho)^{-1} B_{1-\alpha}(k, \rho + 1) \qquad (k = 1, 2, \ldots),$$

where $B_{1-\alpha}(a, b) = \int_0^{1-\alpha} x^{a-1}(1 - x)^{b-1} \, dx$ is the incomplete Beta function, $\rho > 0$, and $\alpha \in (0, 1)$.[6]

In Section 2, we test the goodness-of-fit of the Yule distribution using a parametric bootstrap and powerful test statistics that are not subject to our critique. We do so in three data sets, the one of Chung and Cox (1994) and the two of Giles (2006), both for the parameter $\rho = 1$ chosen by the earlier authors, and two intuitive parameter choices: the maximum likelihood estimator and a moment estimator. The results overwhelmingly reject the Yule distribution. In fact, most p-values are so close to zero that the distribution is rejected at any reasonable significance level. Simple QQ-plots point out where it goes wrong: the Yule distribution seems a fairly accurate approximation of the lower quantiles of the empirical distribution, but puts too much weight in the right tail of the distribution. Consequently, the Yule distribution captures stardom, but not superstardom. On the bright side: the generalized Yule distribution with moment estimators for ρ and α provides an excellent fit in two out of three data sets.

---

[6]It was briefly alluded to by Yule (1924, Sec. III) and Simon (1955, Sec. I), but not derived formally. This is done in the Appendix using a pure birth process with linear birth rates observed at a random time drawn from some finite horizon. Linearity retains the main idea of the snowball effect in [A1]: success breeds success in a proportional way. This derivation differs from the one in Chung and Cox (1994) and the parameters have different meanings. In particular, in the generalized Yule distribution, the constraint that ρ must exceed one as dictated by (2) is absent.

2. Empirical results

2.1. Data

The data description is kept brief. The reader is referred to the original articles for details. The data sets of Giles (2006) are not contained in his article and therefore provided, with his permission, in Table 1.

Our first data set is the one that Chung and Cox (1994) use for their analysis. Their Table I (p. 773) contains, for all Gold Record awardees during $1958 - 1989$, the frequency distribution of performers by their number of Gold Records.

The second data set, used in Giles (2006), measures stardom in terms of the life-length of recordings on the top of the Billboard Hot 100 chart. Table 1 contains, for all recording reaching the number 1 position of the chart during $1955 - 2003$, the frequency distribution of recordings by their number of (not necessarily consecutive) weeks as a number 1 hit.

The third data set, also from Giles (2006), measures success by the amount of number 1 hits of an artist during the same $1955 - 2003$ period. Table 1 contains, for all artists scoring a number 1 hit who released at least 13 recordings into the charts during this period, the frequency distribution of artists by their amount of number 1 hits.

2.2. Analysis

For each data set, we test the statistical validity of:

(a) the discrete Pareto distribution for two parameters: $\rho = 2$ — as it according to (4) has the same tail behavior as the Yule distribution with $\rho = 1$ used by the earlier authors — and the ML-estimator;

(b) the Yule distribution for three parameters: $\rho = 1$, the ML-estimator ρ_{ML}, and a method of moment estimator ρ_{MM} obtained by equating the sample mean $m > 1$ with the expectation $\rho/(\rho - 1)$ and solving for ρ;

(c) the generalized Yule distribution with method of moment estimators ρ_{MM} and α_{MM} obtained by equating the sample's first and second moments with their theoretical counterparts and numerically solving for ρ and α.[7]

We use seven different tests for this purpose:

1. a Chi-square test,

---

[7]ML-estimators are not provided due to numerical problems in the optimization involving the incomplete Beta function.

2. a G-test, also known as likelihood-ratio (LR) test,

3. a discrete Kolmogorov-Smirnov test,

4. a nominal Kolmogorov-Smirnov test,

5. a discrete Cramér-von Mises test,

6. a discrete Watson test,

7. a discrete Anderson-Darling test.

The third and fourth tests are supremum tests, whereas the final three are quadratic tests. See Pettitt and Stephens (1977) and Choulakian et al. (1994) for a more detailed discussion. Rather than relying on tabulated values, we obtain, for each test, p-values by means of a parametric bootstrap with 2,000 bootstrap replications. Consequently, we do not have to worry about the accuracy of asymptotic approximations of finite sample distributions. In particular, our critique [C2], as far as it concerns the required restriction on the expected number of observations in each group, does not apply to this parametric bootstrap version of the Chi-square test.

Log-log rank size plots of the data sets — see Figure 1 for the Gold Records data set — are decidedly nonlinear and all seven goodness-of-fit tests reject the discrete Pareto distribution, regardless of whether one in (3) chooses parameter value $\rho = 2$ or the maximum likelihood estimator.[8] For $\rho = 2$, the Chi-square test has p-value 0.01 in the Hits data; all other p-values (rounded off to two decimals) are 0.00, so the Pareto distribution is rejected at each reasonable significance level.

Table 2 shows that virtually all tests also reject the Yule distribution for the three different parameters. Indeed, in a vast majority of the cases (44 out of the 63 p-values in the table), the p-value (rounded off to two decimals) is 0.00, rejecting the Yule distribution at each reasonable significance level. The Chi-square test has a notable lack of power in comparison with the other tests: it does not reject the Yule distribution with $\rho = 1$ or $\rho = \rho_{ML}$ in the Gold Records data and the Hits data at a 5% significance level. For the Gold Records data, also the G-test and the discrete Watson test cannot reject the Yule distribution with $\rho = 1$ at a 5% significance level, although the latter does reject at a 10% significance level. Similarly, for the Hits data, also the G-test and the discrete Kolmogorov-Smirnov, Cramér-von Mises, and Anderson-Darling tests cannot reject the

---

[8]The ML-estimates are 1.863 (std. dev. 0.024) for the Gold Records data, 1.825 (std. dev. 0.028) for the Weeks data, and 1.956 (std. dev. 0.065) for the Hits data.

Yule distribution with $\rho = 1$ at a 5% significance level, although all these p-values lie around 0.10 and would therefore reject at slightly higher significance levels.

Although the ML-estimator is consistent, efficient, and asymptotically normally distributed, it might suffer from finite sample biases. Therefore, Table 2 also provides the method of moment estimator. Moreover, as a robustness check in connection with finite sample biases, it provides two standard errors for the ML- and MM-estimates. The first is the conventional one based on the asymptotic normality of the estimator. The second is based on a bootstrap with 2,000 runs, each time drawing a new sample of the same size with replacement from the original data and obtaining the corresponding estimate for ρ. The bootstrapped standard error is the sample standard error corresponding to these 2,000 estimates and is slightly smaller than the asymptotic standard error (ML) or of the same magnitude (MM). The standard errors for the moment estimates in Table 3 are bootstrapped standard errors and were obtained analogously.

The test results in Table 3 show that according to most tests, the generalized Yule distribution provides a good fit to the Gold Record and Hits data.

We visualize in two ways why the Yule distribution provides a bad fit to the data. Firstly, the QQ-plot in Figure 1 for the ML-estimate of ρ in the Gold Record data set shows that the right tail of the Yule distribution is too heavy to fit the data well; QQ-plots for the other parameters and data sets exhibit the same and are therefore omitted. In particular, the Yule distribution starts to fail at the 94% quantile in case of the Gold Records data. For the Weeks and the Hits data the fit for $\rho = \rho_{ML}$ becomes bad at the 90% and 93% quantiles, respectively. Secondly, the third graph in Figure 1 clearly illustrates the goodness-of-fit of the generalized Yule distribution: its cumulative distribution function is very close to the empirical distribution function (EDF). The Yule distribution, on the other hand, allocates too much weight to high observations and lies below the empirical distribution in the right tail (which is even more dramatically displayed in the QQ-plot). Hence, the Yule distribution can capture stardom, but fails to model superstardom.

As a more general insight, our analysis shows that (1) the standard Chi-squared test is not always appropriate due to its 'omnibus' character and lack of power (Moore, 1986) and (2) bootstrapping is a good method of obtaining the p-values for statistical tests: one no longer needs to rely on possibly bad asymptotic approximations of finite sample distributions.

8

3. Summary and concluding remarks

Due to the social dimension[9] of enjoying cultural goods like music, using snowball effects as a cause of superstardom is intuitively appealing. Our paper concerns the literature using the Yule distribution, which implements a snowball effect via [A1].

We provided three comments, [C1] to [C3], to the traditional analysis of the Yule distribution for superstar data. In Section 2, we tested the goodness-of-fit of the Yule distribution using a parametric bootstrap and powerful test statistics that are not subject to our comments. The tests overwhelmingly reject the Yule distribution: it is a fairly accurate approximation of the lower quantiles of the empirical distribution, but overestimates the snowball effect that makes consumers purchase records of the most successful artists. In other words, the Yule distribution captures stardom, but not superstardom. The generalized Yule distribution, however, provides an excellent fit to two of the three data sets.

We leave it to future research whether a different model of the snowball effect results in a good fit for all three data sets.

References

Adler, M., 1985. Stardom and talent. American Economic Review 75, 208–212.

Adler, M., 2006. Stardom and talent. In: Ginsburgh, V.A., and Throsby, D. (Eds.), Handbook of the Economics of Art and Culture. Amsterdam: Elsevier, pp. 895–906.

Bourdieu, P., 1989. Distinction: A social critique of the judgment of taste. London: Routledge.

Choulakian, V., Lockhart, R.A., Stephens, M.A., 1994. Cramér-von Mises statistics for discrete distributions. The Canadian Journal of Statistics 22, 125–137.

Chung, K.H., Cox, R.A.K., 1994. A stochastic model of superstardom: an application of the Yule distribution. Review of Economics and Statistics 76, 771–775.

Chung, K.H., Cox, R.A.K., 1998. Consumer behavior and superstardom. Journal of Socio-Economics 27, 263–270.

Consul, P.C., 1991. Evolution of surnames. International Statistical Review 59, 271–278.

Cox, R.A.K., Kleiman, R.T., 2000. A stochastic model of superstardom: evidence from *Institutional Investor*'s All-American Research Team. Review of Financial Economics 9, 43–53.

---

[9]See, for instance, Stigler and Becker (1977) or the French sociologist Bourdieu (1989).

Friedman, M., 1953. The methodology of positive economics. In: Essays in positive economics. Chicago: University of Chicago Press, pp. 3–43.

Giles, D.E., 2006. Superstardom in the US popular music industry revisited. Economics Letters 92, 68–74.

Moore, D.S., 1986. Tests of the Chi-squared type. In: D'Agostino, R.B., and Stephens, M.A. (Eds.), Goodness-of-Fit Techniques. New York: Marcel Dekker, pp. 63–95.

Pettitt, A.N., Stephens, M.A., 1977. The Kolmogorov-Smirnov goodness-of-fit statistic with discrete and grouped data. Technometrics 19, 205–210.

Rosen, S., 1981. The economics of superstars. American Economic Review 71, 845–858.

Simon, H.A., 1955. On a class of skew distribution functions. Biometrika 42, 425–440.

Stigler, G.J., Becker, G.S., 1977. De gustibus non est disputandum. American Economic Review 67, 76–90.

Yule, G.U., 1924. A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, F.R.S. Philosophical Transactions of the Royal Society of London, Series B, 213, 21–87.

Weeks			Hits		
# weeks	# obs.	freq.	# hits	# obs.	freq.
1	337	0.349	1	116	0.473
2	249	0.258	2	57	0.233
3	139	0.144	3	30	0.122
4	93	0.096	4	13	0.053
5	47	0.049	5	10	0.041
6	35	0.036	6	4	0.016
7	21	0.022	7	1	0.004
8	13	0.013	8	1	0.004
9	9	0.009	9	4	0.016
10	8	0.008	10	2	0.008
11	5	0.005	11	1	0.004
12	2	0.002	12	2	0.008
13	2	0.002	13	1	0.004
14	4	0.004	14	1	0.004
15	1	0.001	15	1	0.004
total	965		16	1	0.004
			total	245	

Table 1: Frequency distribution of Weeks and Hits data

This table displays the frequency distribution for the Weeks and the Hits data, both from Giles (2006). The Weeks data measure for all recordings reaching the number 1 position of the the Billboard Hot 100 chart during $1955 - 2003$, their number of (not necessarily consecutive) weeks as a number 1 hit. The Hits data measure success by the number of number 1 hits by an artist during the same $1955 - 2003$ period. It contains, for all artists scoring a number 1 hit who released at least 13 recordings into the charts during this period, their amount of number 1 hits.

Gold records	$\rho = 1$	ML-estimate	MM-estimate
ρ	1.000	1.138 (0.040, 0.032)	1.454 (0.023, 0.023)
Chi-square test	0.18	0.11	0.00
G-test (=LR-test)	0.15	0.02	0.00
Discrete Kolmogorov-Smirnov test	0.02	0.00	0.00
Nominal Kolmogorov-Smirnov test	0.05	0.00	0.00
Discrete Cramér-von Mises test	0.01	0.00	0.00
Discrete Watson test	0.07	0.02	0.00
Discrete Anderson-Darling test	0.01	0.00	0.00
Weeks			
ρ	1.000	1.137 (0.048, 0.030)	1.570 (0.023, 0.023)
Chi-square test	0.00	0.00	0.00
G-test (=LR-test)	0.00	0.00	0.00
Discrete Kolmogorov-Smirnov test	0.00	0.00	0.00
Nominal Kolmogorov-Smirnov test	0.00	0.00	0.00
Discrete Cramér-von Mises test	0.00	0.00	0.00
Discrete Watson test	0.00	0.00	0.00
Discrete Anderson-Darling test	0.00	0.00	0.00
Hits			
ρ	1.000	1.347 (0.118, 0.091)	1.659 (0.075, 0.079)
Chi-square test	0.23	0.17	0.03
G-test (=LR-test)	0.09	0.03	0.00
Discrete Kolmogorov-Smirnov test	0.10	0.00	0.00
Nominal Kolmogorov-Smirnov test	0.02	0.00	0.00
Discrete Cramér-von Mises test	0.10	0.00	0.00
Discrete Watson test	0.00	0.00	0.00
Discrete Anderson-Darling test	0.11	0.00	0.00

Table 2: Goodness-of-fit tests for the Yule distribution

The p-values of the various statistical tests for the three data sets under consideration. These p-values are based on a parametric bootstrap from the null distribution. As a robustness check, two standard errors (in parentheses) are given for the ML- and MM-estimate: first the conventional one based on the asymptotic normality of the estimator, then a bootstrapped one. The method of moment estimator ρ_{MM} is obtained by equating the sample mean with the expectation $\rho/(\rho - 1)$ and solving for ρ.

Gold records	MM-estimates
ρ	0.368 (0.017)
α	0.109 (0.022)
Chisq test	0.01
G-test (=LR-test)	0.03
Discrete Kolmogorov-Smirnov test	0.24
Nominal Kolmogorov Smirnov test	0.05
Discrete Cramer-von Mises test	0.18
Discrete Watson test	0.02
Discrete Anderson-Darling test	0.18
Weeks	
ρ	0.456 (0.006)
α	0.136 (0.007)
Chisq test	0.00
G-test (=LR-test)	0.00
Discrete Kolmogorov-Smirnov test	0.00
Nominal Kolmogorov Smirnov test	0.00
Discrete Cramer-von Mises test	0.00
Discrete Watson test	0.00
Discrete Anderson-Darling test	0.00
Hits	
ρ	0.383 (0.015)
α	0.167 (0.007)
Chisq test	0.16
G-test (=LR-test)	0.09
Discrete Kolmogorov-Smirnov test	0.34
Nominal Kolmogorov Smirnov test	0.09
Discrete Cramer-von Mises test	0.28
Discrete Watson test	0.08
Discrete Anderson-Darling test	0.35

Table 3: Goodness-of-fit tests for the generalized Yule distribution

The p-values of the various statistical tests for the three data sets under consideration. These p-values are based on a parametric bootstrap from the null distribution. In parentheses are the bootstrapped standard errors. The method of moment estimators are obtained by equating the sample's first and second moments with their theoretical counterparts and numerically solving for ρ and α.
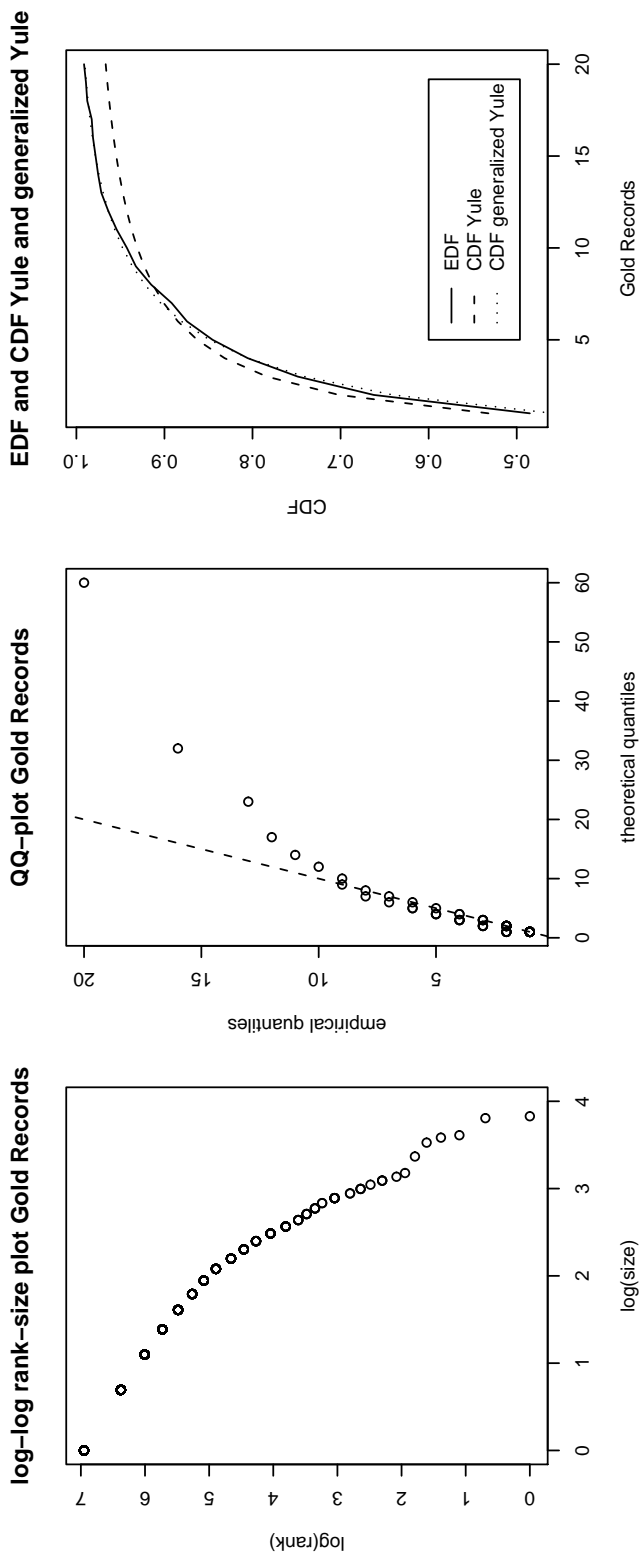
Figure 1: Graphs to evaluate the fit of candidate distributions in the Gold Records data set

The first graph is the log-log rank-size plot. The second graph is a QQ-plot of the empirical quantiles based on the Yule distribution with $\rho = \rho_{ML}$ as a function of the theoretical ones. The diagonal is also given. The third graph contains the empirical distribution function, as well as the cumulative distribution functions of the generalized Yule distribution with its moment estimators and the Yule distribution with $\rho = \rho_{ML}$.

14

APPENDIX. A MODIFICATION OF THE YULE DISTRIBUTION

The two-parameter variant of the Yule distribution discussed below is alluded to in the literature (Yule, 1924, Sec. III; Simon, 1955, Sec. I), but to our knowledge, the derivation below is novel.

DERIVATION: Consider a pure birth process on state space $\mathbb{N}$ with linear birth rates: there is a $\lambda > 0$ such that the birth rate in state $k \in \mathbb{N}$ is $k\lambda$. The initial state is $k = 1$. Formally, this is a continuous time Markov process X on $\mathbb{N}$ with

$$\dot{p} = pQ \qquad \text{and initial state } p(0) = (1, 0, 0, \ldots),$$

where $p = (p_k(t))_{k \in \mathbb{N}, t \geq 0}$ is the vector of probabilities $p_k(t) = \mathbb{P}(X(t) = k)$ of being in state $k \in \mathbb{N}$ at time $t \geq 0$ and, for all $i, j \in \mathbb{N}$:

$$Q_{ij} = \begin{cases} -i\lambda & \text{if } j = i, \\ i\lambda & \text{if } j = i + 1, \\ 0 & \text{otherwise.} \end{cases}$$

In more convenient notation:

$$\begin{aligned} \dot{p}_1 &= -\lambda p_1, \\ \dot{p}_k &= (k-1)\lambda p_{k-1} - k\lambda p_k \qquad (k > 1). \end{aligned}$$

For each $k \in \mathbb{N}$ and $t \geq 0$, the solution, easily verified by substitution above, is

$$p_k(t) = e^{-\lambda t}(1 - e^{-\lambda t})^{k-1}.$$

Suppose the process is observed at a random time $T \geq 0$ (independent from the birth process), drawn from an exponential distribution cut off at time $T_0 > 0$: the distribution of T is an exponential distribution conditional on $T \in [0, T_0]$. Hence, its density is

$$f(t) = \begin{cases} \frac{1}{1-e^{-\mu T_0}} \mu e^{-\mu t} & \text{if } t \in [0, T_0], \\ 0 & \text{otherwise.} \end{cases}$$

Conditioning on the realization of T, the probability of the birth process having reached $k \in \mathbb{N}$ is

$$\begin{aligned} \mathbb{P}(X(T) = k) &= \int_0^{T_0} p_k(t) f(t) \, dt \\ &= \int_0^{T_0} e^{-\lambda t}(1 - e^{-\lambda t})^{k-1} \frac{1}{1 - e^{-\mu T_0}} \mu e^{-\mu t} \, dt \\ &= \int_0^{T_0} e^{-\lambda t}(1 - e^{-\lambda t})^{k-1} \frac{1}{1 - e^{-\mu T_0}} \mu (e^{-\lambda t})^{\mu/\lambda} \, dt. \end{aligned} \qquad \text{(A1)}$$

Use the change of variable $x = 1 - e^{-\lambda t}$. With a common abuse of notation, $dx = \lambda e^{-\lambda t}\, dt$, and the integration bounds become $\lim_{t\to 0}(1-e^{-\lambda t}) = 0$ and $\lim_{t\to T_0}(1-e^{-\lambda t}) = 1-e^{-\lambda T_0}$. Substitution in (A1) gives

$$\mathbb{P}(X(T) = k) = \frac{1}{1 - e^{-\mu T_0}} \cdot \frac{\mu}{\lambda} \int_0^{1-e^{-\lambda T_0}} x^{k-1}(1-x)^{\mu/\lambda}\, dx. \tag{A2}$$

Define $\rho = \mu/\lambda$ and $\alpha = e^{-\lambda T_0}$. Then $\alpha^\rho = e^{-\mu T_0}$, so (A2) becomes

$$p_k = \mathbb{P}(X(T) = k) = \frac{\rho}{1 - \alpha^\rho} \int_0^{1-\alpha} x^{k-1}(1-x)^\rho\, dx = \frac{\rho}{1-\alpha^\rho} B_{1-\alpha}(k, \rho+1), \tag{A3}$$

where for $\varepsilon \in (0,1)$, $B_\varepsilon(\cdot, \cdot)$ is the incomplete Beta function:

$$B_\varepsilon(a, b) = \int_0^\varepsilon x^{a-1}(1-x)^{b-1}\, dx.$$

By choosing λ, μ, T_0 appropriately, the derivation of the two-parameter variant of the Yule distribution allows us to choose $\alpha \in (0,1)$ and $\rho > 0$ arbitrarily.

RELATION WITH THE STANDARD YULE DISTRIBUTION: Fix $\rho > 0$. Then

$$\lim_{\alpha \downarrow 0} \frac{\rho}{1 - \alpha^\rho} B_{1-\alpha}(k, \rho+1) = \rho B(k, \rho+1),$$

i.e., the probability mass function of the traditional Yule distribution is a limiting case of its two-parameter modification. It is obtained by sampling the process above at an exponentially distributed time, without conditioning on some finite horizon.

CUMULATIVE DISTRIBUTION FUNCTION (CDF): As $\sum_{\ell=k+1}^\infty x^{\ell-1} = x^k(1-x)^{-1}$ for each $x \in [0,1)$ and $k \in \mathbb{N}$:

$$\begin{aligned}
\mathbb{P}(X(T) > k) &\overset{(A3)}{=} \sum_{\ell=k+1}^\infty \frac{\rho}{1-\alpha^\rho} \int_0^{1-\alpha} x^{\ell-1}(1-x)^\rho\, dx \\
&= \frac{\rho}{1-\alpha^\rho} \int_0^{1-\alpha} \sum_{\ell=k+1}^\infty x^{\ell-1}(1-x)^\rho\, dx \\
&= \frac{\rho}{1-\alpha^\rho} \int_0^{1-\alpha} x^k(1-x)^{\rho-1}\, dx \\
&= \frac{\rho}{1-\alpha^\rho} B_{1-\alpha}(k+1, \rho).
\end{aligned}$$

Consequently, the cdf of the generalized Yule distribution is

$$\mathbb{P}(X(T) \le k) = 1 - \frac{\rho}{1-\alpha^\rho} B_{1-\alpha}(k+1, \rho) \qquad (k = 1, 2, \ldots).$$

FIRST AND SECOND MOMENTS: Let $\alpha \in (0,1), \rho > 0$. As $\sum_{k=1}^{\infty} x^k = x/(1-x)$ for each $x \in [0,1)$, differentiation gives

$$\sum_{k=1}^{\infty} kx^{k-1} = (1-x)^{-2}. \tag{A4}$$

By definition, the first moment is

$$
\begin{aligned}
\mu = \sum_{k=1}^{\infty} kp_k \; &\overset{(A3)}{=} \; \rho(1-\alpha^\rho)^{-1} \sum_{k=1}^{\infty} \int_0^{1-\alpha} kx^{k-1}(1-x)^\rho \, dx \\
&= \; \rho(1-\alpha^\rho)^{-1} \int_0^{1-\alpha} \sum_{k=1}^{\infty} kx^{k-1}(1-x)^\rho \, dx \\
&\overset{(A4)}{=} \; \rho(1-\alpha^\rho)^{-1} \int_0^{1-\alpha} (1-x)^{\rho-2} \, dx \\
&= \; \begin{cases} -\frac{\ln \alpha}{1-\alpha} & \text{if } \rho = 1, \\[2mm] \frac{\rho(1-\alpha^{\rho-1})}{(\rho-1)(1-\alpha^\rho)} & \text{if } \rho \neq 1. \end{cases}
\end{aligned}
$$

It follows by induction that $f : [0,1) \to \mathbb{R}$ with $f(x) = (x+1)(1-x)^{-3}$ has n-th order derivative $f^{(n)}(x) = (n+1)!(x+n+1)(1-x)^{-(n+3)}$. Hence, its power series expansion is

$$(x+1)(1-x)^{-3} = \sum_{n=0}^{\infty} \frac{f^{(n)}(0)}{n!} x^n = \sum_{n=0}^{\infty} (n+1)^2 x^n = \sum_{k=1}^{\infty} k^2 x^{k-1}. \tag{A5}$$

Using partial integration in the final equation, the second moment is

$$
\begin{aligned}
\mu_2 = \sum_{k=1}^{\infty} k^2 p_k \; &\overset{(A3)}{=} \; \rho(1-\alpha^\rho)^{-1} \sum_{k=1}^{\infty} \int_0^{1-\alpha} k^2 x^{k-1}(1-x)^\rho \, dx \\
&= \; \rho(1-\alpha^\rho)^{-1} \int_0^{1-\alpha} \sum_{k=1}^{\infty} k^2 x^{k-1}(1-x)^\rho \, dx \\
&\overset{(A5)}{=} \; \rho(1-\alpha^\rho)^{-1} \int_0^{1-\alpha} (x+1)(1-x)^{\rho-3} \, dx \\
&= \; \begin{cases} \frac{2(1-\alpha)+\alpha \ln \alpha}{\alpha(1-\alpha)} & \text{if } \rho = 1, \\[3mm] \frac{2(\alpha-1-2\ln \alpha)}{1-\alpha^2} & \text{if } \rho = 2, \\[3mm] \frac{\rho}{1-\alpha^\rho} \cdot \frac{-2\alpha^{\rho-2}(\rho-1)+\alpha^{\rho-1}(\rho-2)+\rho}{(\rho-1)(\rho-2)} & \text{if } \rho \notin \{1,2\}. \end{cases}
\end{aligned}
$$