

Bentley's conjecture on popularity toplist turnover under random copying

Kimmo Eriksson · Fredrik Jansson ·
Jonas Sjöstrand

Dedicated to George Andrews for his 70th birthday

Received: 30 June 2009 / Accepted: 15 October 2010 / Published online: 28 October 2010
© Springer Science+Business Media, LLC 2010

Abstract Bentley et al. studied the turnover rate in popularity toplists in a ‘random copying’ model of cultural evolution. Based on simulations of a model with population size N , list length ℓ and invention rate μ , they conjectured a remarkably simple formula for the turnover rate: $\ell\sqrt{\mu}$. Here we study an overlapping generations version of the random copying model, which can be interpreted as a random walk on the integer partitions of the population size. In this model we show that the conjectured formula, after a slight correction, holds asymptotically.

Keywords Toplists · Random walk · Integer partitions · Moran model · Popularity distribution

Mathematics Subject Classification (2000) Primary 60G50 · 05A17

This research was supported by the CULTAPTATION project (European Commission contract FP6-2004-NEST-PATH-043434) and the Swedish Research Council.

K. Eriksson (✉) · F. Jansson · J. Sjöstrand
School of Communication, Culture and Communication, Mälardalen University, 721 23 Västerås,
Sweden
e-mail: kimmo.eriksson@mdh.se

F. Jansson
e-mail: fredrik.jansson@mdh.se

J. Sjöstrand
e-mail: jonas.sjostrand@mdh.se

K. Eriksson · F. Jansson · J. Sjöstrand
Centre for the Study of Cultural Evolution at Stockholm University, Stockholm, Sweden

1 Introduction

A pervasive phenomenon in modern culture are toplist lists like Top 100 Baby Names or the Billboard Top 200 Pop Chart. Mathematics is no exception; indeed, the present paper was partly inspired by Andrews and Berndt's paper on the *Top Ten Most Fascinating Formulas in Ramanujan's Lost Notebook* reporting the outcome of a popularity vote on these formulas among experts in the field [1]. Andrews and Berndt admit that their list would very likely look different in another week, and point to the corresponding phenomenon in pop charts: "it is the fate of popular songs to lose their popularity and fade off the charts" (p. 19). Here we will be concerned precisely with this phenomenon of turnover of toplist lists. By the *turnover rate* we will mean the number of entries that have gone off the list after a given time.

Our aim is to prove (a modified version of) a remarkable conjecture on turnover rates found by Bentley et al. [3]. These authors analyzed empirical data on the turnover of toplist lists of various things: baby names, pop albums, and dog breeds. Their data allowed them to study varying list lengths (like Top 10, Top 100, etc.), which yielded intriguing results: The turnover rate seemed to be approximately *proportional to the list length* and largely *independent of the underlying population size*.

In order to find a theoretical explanation for this empirical finding, Bentley et al. then simulated cultural evolution of the popularity distribution of cultural variants (say, pop songs) under a simple random copying model. Each individual of a new generation is assumed to copy the favorite song of a randomly drawn individual of the previous generation, but with a small probability μ the individual instead invents a new song. Simulations of this model gave results consistent with the empirical data, and the authors observed the following pattern (without any attempt at analytical verification), which we will refer to as Bentley's conjecture.

Conjecture 1 (Bentley's conjecture [3]) *Under the above random copying model with N individuals per generation, list length ℓ , and invention rate μ , the expected turnover rate of the toplist is very close to*

$$\ell\sqrt{\mu} \tag{1}$$

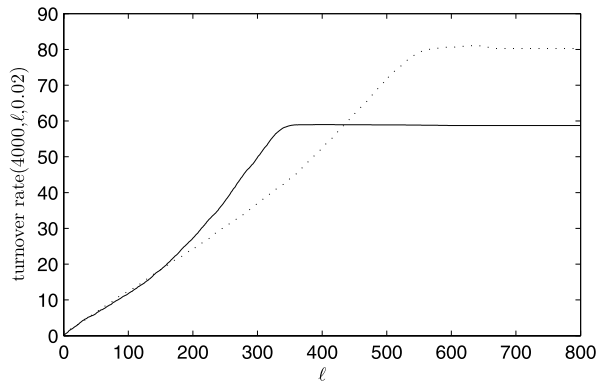
for small μ and sufficiently large N compared to ℓ .

Bentley et al. made no attempt to pin down more precisely the assumptions and the result. Our aim in this paper is to come up with, and prove, a precise formulation of the result. First, we urge the reader to take a moment to appreciate the problem, since it seems to us to be quite novel. As we will discuss at the end, there are a number of well-known stochastic processes that can be interpreted as generating popularity distributions, but we have never before seen considered the question of how often the most popular elements get replaced. If Bentley's conjecture is to be believed, then such questions may have very nice answers.

1.1 Comparison of two random copying models

We will now briefly discuss two extensively studied random copying models from mathematical population genetics, called the *infinite alleles Wright–Fisher model* and the *infinite alleles Moran model*. For all details, we refer to the book by Ewens [6].

Fig. 1 Results of simulations of Bentley's model (*dotted curve*) and the IAM model (*solid curve*), for varying list length ℓ , with fixed population size $N = 4000$ and invention rate $\mu = 0.02$. The simulation first runs for a number of steps so that a stationary distribution is approached. The average turnover rate is then computed as the mean of the turnover rate in the following 200 generations



Bentley's model is equivalent to the infinite alleles Wright–Fisher model. In the present paper we will instead work with the infinite alleles Moran model (henceforth IAM). The IAM model differs from Bentley's model only in the respect that generations are overlapping: Each timestep sees the death of a randomly chosen individual and the birth of a new individual who either inherits the variant of a randomly chosen parent or, with probability μ , invents a new variant.

As described in [6], the two models often give pretty similar results (see also simulations in [8]). The IAM model seems on the whole to be more amenable to exact analysis. For instance, an exact expression for the stationary distribution is known for the IAM, whereas the same expression holds asymptotically for the infinite alleles Wright–Fisher model.

Simulations show that the two models also seem to behave similarly with respect to toplist turnover per generation, if we in the IAM model define a generation as N time steps. Figure 1 shows the turnover rate per generation for varying list length ℓ . Clearly, there are three regimes for both models. In the first regime (for short toplists), the two models give similar turnover rates with what looks like a linear dependence on ℓ . The second regime has a slightly convex shape. In the third regime, the turnover rate is constant.

Both Bentley's conjecture and the asymptotic analysis in the present paper apply to the first regime, which we will look into more carefully shortly. The third regime is trivially explained: When the list length is greater than the number of existing songs, then every newly invented song enters the list. Hence, the expected turnover rate in this regime is $N\mu$ in Bentley's model (in the figure, $N\mu = 80$), whereas it is somewhat lower in the IAM model as a song may cease to exist in the same generation it was invented, thus not contributing to the turnover. The second regime awaits closer investigation.

Figure 2 shows the first regime for the IAM model, with varying list length and four different population sizes. We see that, in line with Bentley's conjecture, the turnover rate is roughly independent of N and roughly linear in ℓ . A least squares fit of the four curves yields the linear expression 0.126ℓ , with $r^2 = 0.997$.

Finally, Bentley's conjecture says that the turnover rate shall be proportional to the square root of the invention rate μ . Figure 3 shows that this indeed seems to be

Fig. 2 Results of simulations of the IAM model, for varying ℓ , four different values of N , and constant $\mu = 0.02$. Average turnover rate is computed as the mean turnover rate in 70 generations

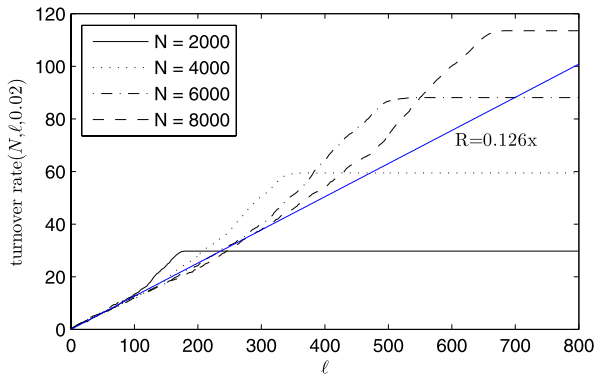
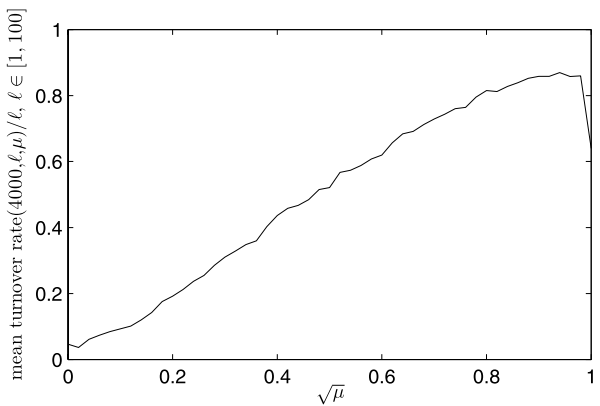


Fig. 3 Results of simulations of the IAM model, with $N = 4000$ and $\mu \in [0, 1]$. The curve shows the mean turnover rate divided by list length, taken over all list lengths $\ell \in \{1, 2, \dots, 100\}$



the case as long as μ is not too close to 1. A least squares fit for the part of the curve where $\sqrt{\mu} \leq 0.8$ yields the linear expression $1.037\sqrt{\mu}$ for the turnover rate, with $r^2 = 0.996$.

1.2 Outline of result and approach

The above simulation results indicate that Bentley’s conjecture applies to the IAM model as well. For this model, Strimling et al. [8] recently obtained an expression for the expected number of variants of popularity k under the stationary distribution. Denoting this number by f_k , they proved that

$$f_k = \mu N \frac{(1 - \mu)^{k-1}}{k} \prod_{i=1}^{k-1} \frac{N - i}{N - i - 1 + i\mu}. \tag{2}$$

This formula will be the starting point for our analysis.

It can be helpful to observe that the popularity distribution of cultural variants can be viewed as an *integer partition* of the population size N . When we talk about a *song* that has k votes, one can think about a particular row of length k of the Young diagram. Thus, the IAM model of cultural evolution can be interpreted as a random

walk on the set of Young diagrams of N squares, where each step involves the death of one random square and the birth of a new square, either through doubling of a random square or creation of a new row of a single square (and then possibly some reordering of the rows so that they are kept in order of decreasing length). In another paper [5] we explicitly use this interpretation to study limit shapes of integer partitions that evolve according to the IMA model (and related stochastic processes). In this paper we will instead use the terminology of songs and votes. The following key concepts will also be important.

- The *popularity* of a song is the number of votes for that song.
- The *toplist* is the set of the ℓ most popular songs. (This is not well-defined if two or more songs have the same popularity, but then we can use any rule to decide which of them should be included on the toplist as it does not matter for our analysis.)
- A *generation* consists of N time steps.
- Let $\alpha > 0$ be some constant. The *turnover rate* is the number of songs that are on the toplist at a given time t_0 but not α generations later. (We will only be interested in the expected turnover rate which is independent of t_0 since we assume a stationary distribution.)

As it turns out, Bentley's conjecture needs to be modified slightly. We will prove the following result:

Theorem 1 *Let $\alpha > 0$ be any constant, suppose N , ℓ , and μ satisfy Assumptions 1, 2 and 3 in Sect. 2, and let ψ be defined as in Assumption 2. Then, under the stationary distribution of the IAM model, the expected turnover rate in α generations (i.e., αN timesteps) is*

$$\sim \sqrt{\psi\alpha/\pi} \cdot \ell \cdot \sqrt{\mu \ln(N/\ell)}.$$

The paper is organized as follows. First we define the notation we will use, and present the assumptions needed for the theorem (Sect. 2). We then derive some basic results (like expectations and variance) about the number of songs of a given popularity (Sect. 3), after which we proceed to examine what popularity it takes to qualify on the toplist (Sect. 4). In the most technical part, we analyze the random process describing how the popularity of a song changes over time (Sect. 5). We then integrate the previous results into a proof of the main theorem (Sect. 6). We conclude by a brief discussion of future directions of research.

2 Notation and assumptions

The input to our problem consists of the three variables N , ℓ and μ and the constant α . To make the notation simple and clear in the following sections, we will think of the variables N , ℓ and μ , and functions of those, as depending on a single free variable ω . Thus, when we write e.g. $N/\ell \rightarrow \infty$ we mean that $N_\omega/\ell_\omega \rightarrow \infty$ when $\omega \rightarrow \infty$. However, the dependence on ω will always be invisible as we will drop the index and write e.g. N instead of N_ω .

We will use conventional ordo notation and the symbols \ll , \sim and \lesssim .

- $A \ll B$ means that $A/B \rightarrow 0$.
- $A \sim B$ means that $A/B \rightarrow 1$.
- $A \lesssim B$ means that, for any $\varepsilon > 0$, eventually $|A/B| < 1 + \varepsilon$.

Our result requires that the variables N , ℓ and μ satisfy three assumptions that restrain their asymptotical behavior.

Assumption 1 $N/\ell \rightarrow \infty$.

Assumption 2 $\psi := 1 - \lim \frac{\ln(\mu^{-1})}{\ln(N/\ell)}$ exists, and $0 < \psi < 1$.

Assumption 3 $\mu\ell/\sqrt{\ln(N/\ell)} \rightarrow \infty$.

These assumptions easily imply the following basic asymptotical properties of our variables:

Corollary 1 *The following holds.*

$$N \rightarrow \infty, \tag{3}$$

$$\ell \rightarrow \infty, \tag{4}$$

$$\mu \ln(N/\ell) \rightarrow 0. \tag{5}$$

Proof We have

$$\frac{\ln((\mu \ln(N/\ell))^{-1})}{\ln(N/\ell)} = \frac{\ln(\mu^{-1})}{\ln(N/\ell)} - \frac{\ln \ln(N/\ell)}{\ln(N/\ell)},$$

which tends to $1 - \psi$ by Assumptions 1 and 2. Thus, $\mu \ln(N/\ell) \rightarrow 0$. Clearly, this implies that $\mu \rightarrow 0$ which together with Assumption 3 yields that $\ell \rightarrow \infty$. Finally, by Assumption 1 we get $N \rightarrow \infty$. □

Finally, we will always assume that the IAM model has reached a stationary distribution (cf. [6, 8]).

3 The number of songs of a given popularity

For $1 \leq m \leq N$, let X_m be the number of songs with popularity m (assuming stationary distribution), which is a random variable. From (2) we already have an exact expression for the expected value $f_m = E(X_m)$. In this section we will examine $\text{Var}(X_m)$ and $x_{m,n} := E(X_{m,n})$, where $X_{m,n}$ denotes the number of ordered pairs (A, B) of distinct songs $A \neq B$ such that A has popularity m and B has popularity n , for $1 \leq m, n \leq N$.

Starting with $x_{m,n}$, first note that

$$X_{m,n} = \begin{cases} X_m X_n & \text{if } m \neq n, \\ X_m(X_m - 1) & \text{if } m = n. \end{cases}$$

Lemma 1

$$x_{m,n} \leq f_m f_n.$$

Proof For any $1 \leq i, m \leq N$, let $P(i \mapsto m)$ denote the probability that a song with popularity i will have popularity m at the next time step. For $1 \leq i, j, m, n \leq N$, define $B_{m,n;i,j}$ and $c_{m,n}$ as follows (when applicable).

$$\begin{aligned} B_{m,n;m,n} &:= P(m \mapsto m) + P(n \mapsto n) - 1 \\ &= 1 - \frac{m+n}{N} - a \frac{m+n}{N-1} + a \frac{m(2m-1) + n(2n-1)}{N(N-1)}, \end{aligned}$$

$$B_{m,n;m+1,n} := P(m+1 \mapsto m) = \frac{m+1}{N} - a \frac{m(m+1)}{N(N-1)},$$

$$B_{m,n;m,n+1} := P(n+1 \mapsto n) = \frac{n+1}{N} - a \frac{n(n+1)}{N(N-1)},$$

$$B_{m,n;m-1,n} := P(m-1 \mapsto m) = a \frac{m-1}{N-1} - a \frac{(m-1)^2}{N(N-1)},$$

$$B_{m,n;m,n-1} := P(n-1 \mapsto n) = a \frac{n-1}{N-1} - a \frac{(n-1)^2}{N(N-1)},$$

$$c_{m,n} := \mu(f_n \delta_{m,1} + f_m \delta_{n,1}).$$

All other $B_{m,n;i,j}$ are set to zero. Here, $a := 1 - \mu$, and $\delta_{m,1}$ is 1 if $m = 1$ and zero otherwise.

Since f_m is the expected value of X_m at the stationary distribution, we must have

$$f_m = \mu \delta_{m,1} + P(m+1 \mapsto m) f_{m+1} + P(m-1 \mapsto m) f_{m-1} + P(m \mapsto m) f_m \tag{6}$$

$$= \mu \delta_{m,1} + B_{m,n;m+1,n} f_{m+1} + B_{m,n;m-1,n} f_{m-1} + P(m \mapsto m) f_m, \tag{7}$$

and, similarly,

$$f_n = \mu \delta_{n,1} + P(n+1 \mapsto n) f_{n+1} + P(n-1 \mapsto n) f_{n-1} + P(n \mapsto n) f_n \tag{8}$$

$$= \mu \delta_{n,1} + B_{m,n;m,n+1} f_{n+1} + B_{m,n;m,n-1} f_{n-1} + P(n \mapsto n) f_n. \tag{9}$$

Multiplying (7) by f_n and (9) by f_m and then adding the resulting equations yields

$$2f_m f_n = c_{m,n} + \left(\sum_{1 \leq i, j \leq N} B_{m,n;i,j} f_i f_j \right) + f_m f_n,$$

and after subtracting $f_m f_n$ we obtain

$$f_m f_n = c_{m,n} + \sum_{1 \leq i, j \leq N} B_{m,n;i,j} f_i f_j. \tag{10}$$

Now, for $1 \leq i, j, m, n \leq N$, define $R_{m,n;i,j}$ and $d_{m,n}$ as follows (when applicable).

$$\begin{aligned}
 R_{m,n;m,n} &:= -2mn \frac{a}{N(N-1)}, \\
 R_{m,n;m+1,n} &:= n(m+1) \frac{a}{N(N-1)}, \\
 R_{m,n;m,n+1} &:= m(n+1) \frac{a}{N(N-1)}, \\
 R_{m,n;m-1,n} &:= n(m-1) \frac{a}{N(N-1)}, \\
 R_{m,n;m,n-1} &:= m(n-1)b \frac{a}{N(N-1)}, \\
 R_{m,n;m-1,n+1} &:= -(m-1)(n+1) \frac{a}{N(N-1)}, \\
 R_{m,n;m+1,n-1} &:= -(n-1)(m+1) \frac{a}{N(N-1)}, \\
 d_{m,n} &:= \frac{\mu}{N} ((nf_n - (n+1)f_{n+1})\delta_{m,1} + (mf_m - (m+1)f_{m+1})\delta_{n,1}).
 \end{aligned}$$

All other $R_{m,n;i,j}$ are set to zero. (Note that $f_{N+1} = 0$ by definition.)

Also, for any $1 \leq i, j, m, n \leq N$ such that $i + j \leq N$, let $P((i, j) \mapsto (m, n))$ denote the probability that two distinct songs with popularity i and j will have popularity m and n , respectively, at the next time step. It is not difficult to check that, for any $1 \leq i, j, m, n \leq N$ such that $i + j \leq N$, in fact

$$B_{m,n;i,j} - R_{m,n;i,j} = P((i, j) \mapsto (m, n)). \tag{11}$$

Now, let us drop the assumption of stationary distribution for a while and instead see what happens if we start with one single song with popularity N at time 0. Let $x_{m,n}^{(t)}$ denote the expected number of ordered pairs (A, B) of songs at time t such that A has popularity m and B has popularity n . It follows from (11) that

$$x_{m,n}^{(t+1)} = c_{m,n} - d_{m,n} + \sum_{1 \leq i, j \leq N} (B_{m,n;i,j} - R_{m,n;i,j})x_{i,j}^{(t)}. \tag{12}$$

Let $r_{m,n}^{(t)} := f_m f_n - x_{m,n}^{(t)}$. Subtracting (12) from (10) yields

$$r_{m,n}^{(t+1)} = d_{m,n} + \sum_{1 \leq i, j \leq N} R_{m,n;i,j} f_i f_j + \sum_{1 \leq i, j \leq N} (B_{m,n;i,j} - R_{m,n;i,j})r_{i,j}^{(t)}. \tag{13}$$

At time 0 there is only one song, so $x_{m,n}^{(0)} = 0$ and thus $r_{m,n}^{(0)} = f_m f_n > 0$ for all m and n . Suppose $r_{i,j}^{(t)} \geq 0$ and let us show that $r_{i,j}^{(t+1)} \geq 0$.

First, note that $r_{m,n}^{(t+1)} = f_m f_n > 0$ automatically when $m + n > N$, so in the following we will assume that $m + n \leq N$. If $m + n$ is strictly less than N , $B_{m,n;i,j} - R_{m,n;i,j}$ either vanishes or equals $P((i, j) \mapsto (m, n)) \geq 0$. When $m + n = N$ it can

easily be checked that $B_{m,n;i,j} - R_{m,n;i,j} \geq 0$ assuming that $\mu N \geq 1$ which follows from Assumptions 1 and 2. Thus, the last sum in (13) is nonnegative. Using again that $\mu N \geq 1$, we have $kf_k > (k + 1)f_{k+1}$ for all k , which implies that $d_{m,n} \geq 0$. It remains to examine the sum $\sum_{1 \leq i,j \leq N} R_{m,n;i,j} f_i f_j$. But this sum can be written as

$$\sum_{1 \leq i,j \leq N} R_{m,n;i,j} f_i f_j = \frac{a}{N(N-1)} \left[((m-1)f_{m-1} - mf_m)(nf_n - (n+1)f_{n+1}) + ((n-1)f_{n-1} - nf_n)(mf_m - (m+1)f_{m+1}) \right], \tag{14}$$

which is nonnegative if $m, n > 1$. If $m = 1$ or $n = 1$, then we must check that $d_{m,n} + \sum_{1 \leq i,j \leq N} R_{m,n;i,j} f_i f_j \geq 0$, and this is true if $\mu N \geq 1$.

Thus, we have proved by induction that $r_{m,n}^{(t)} \geq 0$ for all $1 \leq m, n \leq N$ and all $t \geq 0$. Since the process approaches the unique stationary distribution, it follows that $f_m f_n - x_{m,n} \geq 0$ for all m and n . \square

Proposition 1

$$\text{Var}(X_m) \leq f_m.$$

Proof By Lemma 1 we have

$$\begin{aligned} \text{Var}(X_m) &= E(X_m^2) - E(X_m)^2 = E(X_m(X_m - 1)) + E(X_m) - E(X_m)^2 \\ &\leq f_m^2 + E(X_m) - E(X_m)^2 = f_m. \end{aligned} \tag{15}$$

Proposition 2

$$\text{Var}\left(\sum_{k=m}^N X_k\right) \leq \sum_{k=m}^N f_k.$$

Proof

$$\text{Var}\left(\sum_{k=m}^N X_k\right) = \sum_{m \leq i,j \leq N} (E(X_i X_j) - E(X_i)E(X_j)) \tag{16}$$

$$= \sum_{m \leq i,j \leq N} (x_{i,j} - f_i f_j) + \sum_{k=m}^N f_k \tag{17}$$

$$\leq \sum_{k=m}^N f_k, \tag{17}$$

where the last inequality follows from Lemma 1. \square

4 What does it take to qualify on the toplist?

In this section we will analyze the asymptotics of the number of votes that are needed to qualify on the toplist. To this end we will shortly define and examine two related variables, \hat{K} and \hat{k} . First, we need a couple of lemmas.

Lemma 2 For any m ,

$$f_m \lesssim \mu N(1 - \mu)^m / m,$$

and for $m = O(\frac{\ln(N/\ell)}{\mu})$,

$$f_m \sim \mu N(1 - \mu)^m / m.$$

Proof By (2),

$$f_m = \mu N \frac{(1 - \mu)^{m-1}}{m} \prod_{i=1}^{m-1} \frac{N - i}{N - 1 - (1 - \mu)i}$$

so it suffices to show that the product

$$H := \prod_{i=1}^{m-1} \frac{N - i}{N - 1 - (1 - \mu)i}$$

is asymptotically smaller than 1, and that it tends to 1 if $m = O(\frac{\ln(N/\ell)}{\mu})$. It is easily verified that $\frac{N-i}{N-1-(1-\mu)i}$ is a decreasing function of i and that it is smaller than 1 if $i > 1/\mu$. Thus we obtain upper and lower bounds for H as follows.

$$H \leq \left(\frac{N - 1}{N - 2 + \mu}\right)^{\lceil 1/\mu \rceil} = \left(1 + \frac{2 - \mu}{N - 2 + \mu}\right)^{\lceil 1/\mu \rceil} \sim \exp\left(\frac{2}{\mu N}\right) \rightarrow 1$$

by Assumptions 1 and 3, and

$$\begin{aligned} H &\geq \left(\frac{N - (m - 1)}{N - 1 - (1 - \mu)(m - 1)}\right)^{m-1} = \left(1 + \frac{1 - \mu(m - 1)}{N - 1 - (1 - \mu)(m - 1)}\right)^{m-1} \\ &\sim \exp\left(\frac{(1 - \mu(m - 1))(m - 1)}{N - 1 - (1 - \mu)(m - 1)}\right), \end{aligned}$$

which tends to 1 if $m = O(\frac{\ln(N/\ell)}{\mu})$ since

$$\frac{(\ln(N/\ell))^2}{\mu N} = \frac{1}{\mu \ell} \cdot \frac{(\ln(N/\ell))^2}{N/\ell} = o(\mu^{-1} \ell^{-1}) \rightarrow 0$$

by Assumption 3. □

For $1 \leq m \leq N$, let $S_m = \sum_{k=m}^N X_k$ be the number of songs with popularity at least m .

Lemma 3 *If $\mu^{-1} \ll m = O(\frac{\ln(N/\ell)}{\mu})$ then*

$$E(S_m) = \sum_{k=m}^N f_k \sim f_m/\mu.$$

Proof Let ϕ be an integer variable such that $\mu^{-1} \ll \phi \ll m$. First we divide the sum into two terms:

$$\sum_{k=m}^N f_k = f_m \left(\overbrace{\sum_{k=m}^{m+\phi} f_k/f_m}^A + \overbrace{\sum_{k=m+\phi+1}^N f_k/f_m}^B \right).$$

By Lemma 2 and the assumption on ϕ , we have

$$A \sim \sum_{k=m}^{m+\phi} \frac{m}{k} (1-\mu)^{k-m} \sim \sum_{k=m}^{m+\phi} (1-\mu)^{k-m} = \frac{1 - (1-\mu)^{\phi+1}}{\mu} \sim \frac{1 - e^{-\phi\mu}}{\mu} \sim \frac{1}{\mu}.$$

Once again by Lemma 2 and the assumption on ϕ , we have

$$\begin{aligned} B &\lesssim \sum_{k=m+\phi+1}^N \frac{m}{k} (1-\mu)^{k-m} \\ &< \sum_{k=m+\phi+1}^{\infty} (1-\mu)^{k-m} = \frac{(1-\mu)^{\phi+1}}{\mu} \sim \frac{e^{-\phi\mu}}{\mu} = o(1/\mu). \end{aligned} \tag{18}$$

□

Let \hat{K} be the popularity of the ℓ th most popular song, i.e. \hat{K} is the largest integer such that $S_{\hat{K}} \geq \ell$. In other words, \hat{K} is the number of votes needed to qualify on the toplist. In order to estimate \hat{K} we will study the related measure \hat{k} , defined as the largest integer such that $E(S_{\hat{k}}) \geq \ell$. Below we first determine the asymptotics of \hat{k} (Proposition 3). We will then compute the probability of a large deviation of \hat{K} from \hat{k} (Proposition 5).

Proposition 3 *We have*

$$\hat{k} \sim \psi \frac{\ln(N/\ell)}{\mu}.$$

Proof By Lemma 3, it suffices to show that $f_{(\psi+\varepsilon)\mu^{-1}\ln(N/\ell)} \ll \mu\ell$ and $f_{(\psi-\varepsilon)\mu^{-1}\ln(N/\ell)} \gg \mu\ell$ for any (sufficiently small) fixed $\varepsilon > 0$.

By Lemma 2, we have

$$\frac{f_{(\psi\pm\varepsilon)\mu^{-1}\ln(N/\ell)}}{\mu\ell} \sim \frac{1}{\mu\ell} \frac{\mu N (1-\mu)^{(\psi\pm\varepsilon)\mu^{-1}\ln(N/\ell)}}{(\psi\pm\varepsilon)\mu^{-1}\ln(N/\ell)} =: A_{\pm\varepsilon}.$$

Taking the logarithm yields

$$\begin{aligned}
 \ln A_{\pm\varepsilon} &= -\ln(\mu^{-1}) + \ln(N/\ell) + (\psi \pm \varepsilon)\mu^{-1} \ln(N/\ell) \ln(1 - \mu) \\
 &\quad - \ln(\psi \pm \varepsilon) - \ln \ln(N/\ell) \\
 &= -\ln(\mu^{-1}) + \ln(N/\ell) - (1 + o(1))(\psi \pm \varepsilon) \ln(N/\ell) \\
 &\quad - \ln(\psi \pm \varepsilon) - \ln \ln(N/\ell) \\
 &= \{\text{Assumption 2}\} \\
 &= -(1 - \psi + o(1)) \ln(N/\ell) + \ln(N/\ell) \\
 &\quad - (1 + o(1))(\psi \pm \varepsilon) \ln(N/\ell) - \ln(\psi \pm \varepsilon) - \ln \ln(N/\ell) \\
 &= (\mp\varepsilon + o(1)) \ln(N/\ell) - \ln(\psi \pm \varepsilon) - \ln \ln(N/\ell) \rightarrow \mp\infty.
 \end{aligned}$$

Thus, $A_{+\varepsilon}$ tends to zero and $A_{-\varepsilon}$ tends to infinity. □

Proposition 4

$$f_{\hat{k}+o(\mu^{-1})} \sim \mu\ell.$$

Proof From Lemma 3 and the definition of \hat{k} it follows that $f_{\hat{k}} \sim \mu\ell$. Then Lemma 2 and Proposition 3 tell us that

$$f_{\hat{k}+o(\mu^{-1})} \sim f_{\hat{k}} \frac{\hat{k}}{\hat{k} + o(\mu^{-1})} (1 - \mu)^{o(\mu^{-1})} \sim f_{\hat{k}}. \quad \square$$

Proposition 5 Suppose $0 < \rho = o(\mu^{-1})$. Then the following holds:

$$P(|\hat{K} - \hat{k}| > \rho) \lesssim 2\rho^{-2}\mu^{-2}\ell^{-1}.$$

Proof Without loss of generality, we can assume that ρ is an integer. It follows from the definition of \hat{K} that

$$P(\hat{K} - \hat{k} > \rho) \leq P(S_{\hat{k}+\rho} > \ell)$$

and

$$P(\hat{k} - \hat{K} > \rho) \leq P(S_{\hat{k}-\rho} < \ell).$$

Due to the assumption on ρ , Proposition 4 tells us that

$$E(S_{\hat{k}\pm\rho}) = E(S_{\hat{k}}) \mp \sum_{i=1}^{\rho} f_{\hat{k}\pm i} \sim \ell \mp \rho\mu\ell,$$

and we also have

$$\begin{aligned} \text{Var}(S_{\hat{k}\pm\rho}) &\leq \{\text{Proposition 2}\} \leq E(S_{\hat{k}\pm\rho}) \\ &\sim \{\text{Lemma 3}\} \sim f_{\hat{k}\pm\rho}/\mu \sim \{\text{Proposition 4}\} \sim \ell. \end{aligned}$$

Finally, combining the observations above, we obtain

$$\begin{aligned} P(S_{\hat{k}\pm\rho} \geq \ell) &\leq P(|S_{\hat{k}\pm\rho} - E(S_{\hat{k}\pm\rho})| > |\ell - E(S_{\hat{k}\pm\rho})|) \\ &\leq \{\text{Chebyshev’s inequality}\} \\ &\leq \frac{\text{Var}(S_{\hat{k}\pm\rho})}{|\ell - E(S_{\hat{k}\pm\rho})|^2} \\ &\sim \frac{\ell}{(\rho\mu\ell)^2}. \end{aligned} \quad \square$$

5 How the popularity of a song changes over time

Define the probabilities

$$p^{\text{left}}(k) := \frac{k}{N} \left(1 - (1 - \mu) \frac{k - 1}{N - 1} \right)$$

and

$$p^{\text{right}}(k) := \left(1 - \frac{k}{N} \right) (1 - \mu) \frac{k}{N - 1}$$

and $p^{\text{stay}}(k) := 1 - p^{\text{left}}(k) - p^{\text{right}}(k)$. It follows from the definition of the IAM model that $p^{\text{left}}(k)$ and $p^{\text{right}}(k)$ are the probabilities for a song with k votes to lose and gain a vote, respectively, in the next time step; $p^{\text{stay}}(k)$ is the probability that the number of votes for this song is not affected in this step.

Given a positive integer $\bar{k} \leq N$, define a random integer sequence $(K_t^{(\bar{k})})_{t=0}^\infty$ as follows. Put $K_0^{(\bar{k})} := \bar{k}$ and, assuming that $K_0^{(\bar{k})}, K_1^{(\bar{k})}, \dots, K_{t-1}^{(\bar{k})}$ have already been defined, let $K_t^{(\bar{k})} := K_{t-1}^{(\bar{k})} - 1$ with probability $p^{\text{left}}(K_{t-1}^{(\bar{k})})$ and $K_t^{(\bar{k})} := K_{t-1}^{(\bar{k})} + 1$ with probability $p^{\text{right}}(K_{t-1}^{(\bar{k})})$; otherwise, $K_t^{(\bar{k})} := K_{t-1}^{(\bar{k})}$. Thus, the sequence $(K_t^{(\bar{k})})_{t=0}^\infty$ describes the evolution of the popularity of a song that has \bar{k} votes from the beginning.

We will be interested in assessing the evolution of the random process after α generations, that is, αN time steps. In order to get a grip on this, we will define and examine three other random integer sequences derived from $(K_t^{(\bar{k})})_{t=0}^\infty$. We begin by a brief overview, saving the details until later. For $t = 1, 2, \dots$:

- $U_t^{(\bar{k})}$ is basically $K_t^{(\bar{k})} - K_{t-1}^{(\bar{k})}$, but adjusted in such a way that $P(U_t^{(\bar{k})} = -1) = p^{\text{left}}(\bar{k})$ and $P(U_t^{(\bar{k})} = 1) = p^{\text{right}}(\bar{k})$;

- $V_t^{(\bar{k})}$ is the said adjustment, i.e., $V_t^{(\bar{k})} := K_t^{(\bar{k})} - K_{t-1}^{(\bar{k})} - U_t^{(\bar{k})}$;
- $V_t^{(\bar{k},\delta)}$ is basically $V_t^{(\bar{k})}$, but adjusted to zero if $|K_{t-1}^{(\bar{k})} - \bar{k}| > \delta$.

For notational convenience, we define symbols for the sums of the first αN elements of these sequences: $\tilde{U}^{(\bar{k})} := \sum_{t=1}^{\alpha N} U_t^{(\bar{k})}$, $\tilde{V}^{(\bar{k})} := \sum_{t=1}^{\alpha N} V_t^{(\bar{k})}$ and $\tilde{V}^{(\bar{k},\delta)} := \sum_{t=1}^{\alpha N} V_t^{(\bar{k},\delta)}$.

With this notation, we can express the total change over α generations to the popularity of a song that starts with \bar{k} votes as

$$K_{\alpha N}^{(\bar{k})} - \bar{k} = \tilde{U}^{(\bar{k})} + \tilde{V}^{(\bar{k})}.$$

The purpose of this is that $\tilde{U}^{(\bar{k})}$ approximates the total change by assuming that the probabilities for going left and right are constant, and $\tilde{V}^{(\bar{k})}$ is the total adjustment one must make to that approximation.

Our aim in this section will be to prove that it is unlikely that the total change is large (Proposition 6), and unlikely that the total adjustment is large (Proposition 7). To achieve this, we must first examine the sequences and then their sums.

5.1 The $U_t^{(\bar{k})}$ sequence

Define the random integer sequence $(U_t^{(\bar{k})})_{t=1}^{\infty}$ as follows.

- If $K_t^{(\bar{k})} < K_{t-1}^{(\bar{k})}$, then we put $U_t^{(\bar{k})} = -1$ with probability $\min\{1, p^{\text{left}}(\bar{k})/p^{\text{left}}(K_{t-1}^{(\bar{k})})\}$, otherwise $U_t^{(\bar{k})} = 0$.
- If $K_t^{(\bar{k})} > K_{t-1}^{(\bar{k})}$, then we put $U_t^{(\bar{k})} = 1$ with probability $\min\{1, p^{\text{right}}(\bar{k})/p^{\text{right}}(K_{t-1}^{(\bar{k})})\}$, otherwise $U_t^{(\bar{k})} = 0$.
- If $K_t^{(\bar{k})} = K_{t-1}^{(\bar{k})}$, then we put $U_t^{(\bar{k})} = -1$ with probability

$$\max\left\{0, \frac{p^{\text{left}}(\bar{k}) - p^{\text{left}}(K_{t-1}^{(\bar{k})})}{p^{\text{stay}}(K_{t-1}^{(\bar{k})})}\right\},$$

$U_t^{(\bar{k})} = 1$ with probability

$$\max\left\{0, \frac{p^{\text{right}}(\bar{k}) - p^{\text{right}}(K_{t-1}^{(\bar{k})})}{p^{\text{stay}}(K_{t-1}^{(\bar{k})})}\right\},$$

and $U_t^{(\bar{k})} = 0$ otherwise.

In this way, $U_1^{(\bar{k})}, U_2^{(\bar{k})}, \dots$ become independent identically distributed random variables, with $P(U_t^{(\bar{k})} = -1) = p^{\text{left}}(\bar{k})$ and $P(U_t^{(\bar{k})} = 1) = p^{\text{right}}(\bar{k})$.

5.2 The $V_t^{(\bar{k})}$ sequence

Define the random integer sequence $(V_t^{(\bar{k})})_{t=1}^\infty$ by $V_t^{(\bar{k})} := K_t^{(\bar{k})} - K_{t-1}^{(\bar{k})} - U_t^{(\bar{k})}$. Observe that

$$\begin{aligned}
 &P(V_t^{(\bar{k})} = -1 \mid K_{t-1}^{(\bar{k})} = k) \\
 &= \max\{0, p^{\text{left}}(k) - p^{\text{left}}(\bar{k})\} + \max\{0, p^{\text{right}}(\bar{k}) - p^{\text{right}}(k)\} \\
 &= \max\left\{0, \frac{k - \bar{k}}{N} \left(1 - (1 - \mu) \frac{\bar{k} + k - 1}{N - 1}\right)\right\} \\
 &\quad + \max\left\{0, \left(1 - \frac{\bar{k} + k}{N}\right) (1 - \mu) \frac{\bar{k} - k}{N - 1}\right\} \\
 &\leq 2|k - \bar{k}|/N
 \end{aligned} \tag{19}$$

and

$$\begin{aligned}
 &P(V_t^{(\bar{k})} = +1 \mid K_{t-1}^{(\bar{k})} = k) \\
 &= \max\{0, p^{\text{right}}(k) - p^{\text{right}}(\bar{k})\} + \max\{0, p^{\text{left}}(\bar{k}) - p^{\text{left}}(k)\} \\
 &= \max\left\{0, \left(1 - \frac{\bar{k} + k}{N}\right) (1 - \mu) \frac{k - \bar{k}}{N - 1}\right\} \\
 &\quad + \max\left\{0, \frac{\bar{k} - k}{N} \left(1 - (1 - \mu) \frac{\bar{k} + k - 1}{N - 1}\right)\right\} \\
 &\leq 2|k - \bar{k}|/N.
 \end{aligned} \tag{20}$$

We will also need that

$$\begin{aligned}
 &P(V_t^{(\bar{k})} = +1 \mid K_{t-1}^{(\bar{k})} = k) - P(V_t^{(\bar{k})} = -1 \mid K_{t-1}^{(\bar{k})} = k) \\
 &= (p^{\text{right}}(k) - p^{\text{right}}(\bar{k})) - (p^{\text{left}}(k) - p^{\text{left}}(\bar{k})) = \mu(\hat{k} - k)/N.
 \end{aligned} \tag{21}$$

5.3 The $V_t^{(\bar{k}, \delta)}$ sequence

For any positive integer δ , define the random integer sequence $(V_t^{(\bar{k}, \delta)})_{t=1}^\infty$ by $V_t^{(\bar{k}, \delta)} = V_t^{(\bar{k})}$ if $|K_{t-1}^{(\bar{k})} - \bar{k}| \leq \delta$ and $V_t^{(\bar{k}, \delta)} = 0$ otherwise. From (19), (20) and (21) it follows that

$$P(V_t^{(\bar{k}, \delta)} = -1) \leq 2\delta/N, \tag{22}$$

$$P(V_t^{(\bar{k}, \delta)} = +1) \leq 2\delta/N, \tag{23}$$

$$|E(V_t^{(\bar{k}, \delta)})| \leq \mu\delta/N, \tag{24}$$

$$|E((V_t^{(\bar{k}, \delta)})^2)| \leq 4\delta/N. \tag{25}$$

5.4 Analysis of $\tilde{V}^{(\bar{k}, \delta)}$

The purpose of defining $\tilde{V}^{(\bar{k}, \delta)}$ is that it is easier to analyze than $\tilde{V}^{(\bar{k})}$. Before showing that it is unlikely that $\tilde{V}^{(\bar{k}, \delta)}$ is large, we shall find bounds for its first and second momentum.

From (24), we obtain

$$|E(\tilde{V}^{(\bar{k}, \delta)})| \leq \sum_{t=1}^{\alpha N} |E(V_t^{(\bar{k}, \delta)})| \leq \alpha\mu\delta. \tag{26}$$

It follows from (21) that $|E(V_t^{(\bar{k}, \delta)} \mid \text{“event”})| \leq \mu\delta/N$ for any “event” that takes place at an earlier point in time than t . Using this and (22), (23) and (24), we conclude that, if $1 \leq s < t$,

$$\begin{aligned} |E(V_s^{(\bar{k}, \delta)} V_t^{(\bar{k}, \delta)})| &\leq P(V_s^{(\bar{k}, \delta)} = 1) \cdot |E(V_t^{(\bar{k}, \delta)} \mid V_s^{(\bar{k}, \delta)} = 1)| \\ &\quad + P(V_s^{(\bar{k}, \delta)} = -1) \cdot |E(V_t^{(\bar{k}, \delta)} \mid V_s^{(\bar{k}, \delta)} = -1)| \\ &\leq \frac{2\delta}{N} \cdot \frac{\mu\delta}{N} + \frac{2\delta}{N} \cdot \frac{\mu\delta}{N} \\ &= 4\mu\delta^2/N^2. \end{aligned}$$

Combining this with (25), we obtain

$$E((\tilde{V}^{(\bar{k}, \delta)})^2) = \sum_{t=1}^{\alpha N} E((V_t^{(\bar{k}, \delta)})^2) + 2 \sum_{1 \leq s < t \leq \alpha N} E(V_s^{(\bar{k}, \delta)} V_t^{(\bar{k}, \delta)}) \tag{27}$$

$$\leq 4\alpha\delta + 4\alpha^2\delta^2\mu. \tag{28}$$

Lemma 4 *Suppose $\bar{k} = O(\hat{k})$ and $0 < \delta = O(\hat{k})$. Then, the following holds for $d > 0$:*

$$P(|\tilde{V}^{(\bar{k}, \delta)}| > d) = O(\delta d^{-2}(1 + \delta\mu)).$$

Proof If $d = O(\mu\delta)$, then we would have $\delta d^{-2}(1 + \delta\mu) \rightarrow \infty$ by the assumption on δ together with Corollary 1. Thus, without loss of generality, we may assume that $d \gg \mu\delta$ and hence $d > |E(\tilde{V}^{(\bar{k}, \delta)})|$ by (26). Then the following holds.

$$P(|\tilde{V}^{(\bar{k}, \delta)}| > d) \leq P(|\tilde{V}^{(\bar{k}, \delta)} - E(\tilde{V}^{(\bar{k}, \delta)})| > d - E(\tilde{V}^{(\bar{k}, \delta)})) \tag{29}$$

$$\leq \{\text{Chebyshev’s inequality}\} \leq \frac{\text{Var}(\tilde{V}^{(\bar{k}, \delta)})}{(d - E(\tilde{V}^{(\bar{k}, \delta)}))^2} \tag{30}$$

$$\leq \frac{E((\tilde{V}(\bar{k}, \delta))^2)}{(d - E(\tilde{V}(\bar{k}, \delta)))^2} = \{(26) \text{ and } (28)\} \tag{31}$$

$$= \frac{O(\delta + \delta^2 \mu)}{(d - O(\mu \delta))^2} = \{\text{the assumption that } \delta = O(\hat{k})\} \tag{32}$$

$$= O(\delta d^{-2}(1 + \delta \mu)). \tag{33}$$

□

5.5 Analysis of $\tilde{U}(\bar{k})$

It is much easier to deal with the expected value and variance of $\tilde{U}(\bar{k})$. Since

$$E(U_t^{(\bar{k})}) = p^{\text{right}}(\bar{k}) - p^{\text{left}}(\bar{k}) = -\mu \bar{k} / N \tag{34}$$

and

$$\text{Var}(U_t^{(\bar{k})}) = E((U_t^{(\bar{k})})^2) - E(U_t^{(\bar{k})})^2 = p^{\text{right}}(\bar{k}) + p^{\text{left}}(\bar{k}) - \frac{\mu^2 \bar{k}^2}{N^2} \tag{35}$$

$$= \frac{\bar{k}}{N} \left(2 - \mu - 2(1 - \mu) \frac{\bar{k} - 1}{N - 1} \right) - \frac{\mu^2 \bar{k}^2}{N^2} < \frac{2\bar{k}}{N}, \tag{36}$$

we get

$$E(\tilde{U}(\bar{k})) = \alpha N E(U_t^{(\bar{k})}) = -\alpha \mu \bar{k} \tag{37}$$

and, since $U_1^{(\bar{k})}, U_2^{(\bar{k})}, \dots$ are independent,

$$\text{Var}(\tilde{U}(\bar{k})) = \alpha N \text{Var}(U_t^{(\bar{k})}) < 2\alpha \bar{k}. \tag{38}$$

Lemma 5 *If $\bar{k} = O(\hat{k})$, then, for $d > 0$,*

$$P(|\tilde{U}(\bar{k})| > d) = O(d^{-2} \hat{k}).$$

Proof If $d = O(\mu \hat{k})$ we would have $d^{-2} \hat{k} \rightarrow \infty$ by Corollary 1. Thus, without loss of generality, we may assume that $d \gg \mu \hat{k}$ and hence $d > |E(\tilde{U}(\bar{k}))|$ by (37). Then the following holds.

$$P(|\tilde{U}(\bar{k})| > d) \leq P(|\tilde{U}(\bar{k}) - E(\tilde{U}(\bar{k}))| > d - E(\tilde{U}(\bar{k}))) \tag{39}$$

$$\leq \{\text{Chebyshev’s inequality}\} \leq \frac{\text{Var}(\tilde{U}(\bar{k}))}{(d - E(\tilde{U}(\bar{k})))^2} \tag{40}$$

$$\leq \frac{E((\tilde{U}(\bar{k}))^2)}{(d - E(\tilde{U}(\bar{k})))^2} = \{(37) \text{ and } (38)\} \tag{41}$$

$$= \frac{O(\bar{k})}{(d - \alpha\mu\bar{k})^2} = O(d^{-2}\hat{k}). \tag{42}$$

□

5.6 Analysis of $\tilde{V}(\bar{k})$ and $\tilde{U}(\bar{k}) + \tilde{V}(\bar{k})$

In order to prove our desired propositions about the improbability of large values of $\tilde{V}(\bar{k})$ and $\tilde{U}(\bar{k}) + \tilde{V}(\bar{k})$, we need two additional lemmas. First, we study $\tilde{U}(\bar{k}) + \tilde{V}(\bar{k},\delta)$ instead.

Lemma 6 *If $\bar{k} = O(\hat{k})$ and $0 < \delta = O(\hat{k})$, then the following holds:*

$$P(|\tilde{U}(\bar{k}) + \tilde{V}(\bar{k},\delta)| > \delta/2) = O(\delta^{-2}\hat{k} + \mu).$$

Proof Obviously,

$$P(|\tilde{U}(\bar{k}) + \tilde{V}(\bar{k},\delta)| > \delta/2) \leq P(|\tilde{U}(\bar{k})| > \delta/4) + P(|\tilde{V}(\bar{k},\delta)| > \delta/4).$$

The first term, $P(|\tilde{U}(\bar{k})| > \delta/4)$, is $O(\delta^{-2}\hat{k})$ by Lemma 5. The other term, $P(|\tilde{V}(\bar{k},\delta)| > \delta/4)$, is $O(\delta^{-1} + \mu)$ by Lemma 4 with $d = \delta/4$, and $O(\delta^{-1} + \mu)$ is $O(\delta^{-2}\hat{k} + \mu)$ by the assumption that $\delta = O(\hat{k})$. □

Next we must show that it is unlikely that $\tilde{V}(\bar{k},\delta)$ differs from $\tilde{V}(\bar{k})$.

Lemma 7 *If $\bar{k} = O(\hat{k})$ and $0 < \delta = O(\hat{k})$, then*

$$P(\exists t \leq \alpha N : V_t^{(\bar{k},\delta)} \neq V_t^{(\bar{k})}) = O(\delta^{-2}\hat{k} + \mu).$$

Proof If there is a $t \leq \alpha N$ such that $V_t^{(\bar{k},\delta)} \neq V_t^{(\bar{k})}$, then clearly $|K_{t-1} - \bar{k}| > \delta$ and there must exist a $T \leq \alpha N$ such that $|\sum_{\tau=1}^T (U_\tau^{(\bar{k})} + V_\tau^{(\bar{k},\delta)})| = \lfloor \delta + 1 \rfloor$. Then, by Lemma 6, the probability that $|\sum_{\tau=T+1}^{\alpha N} (U_\tau^{(\bar{k})} + V_\tau^{(\bar{k},\delta)})| \leq \delta/2$ is $1 - O(\delta^{-2}\hat{k} + \mu)$. But if this probable event happens, then we must have $|\sum_{\tau=1}^{\alpha N} (U_\tau^{(\bar{k})} + V_\tau^{(\bar{k},\delta)})| > \delta/2$ which, again by Lemma 6, happens with probability $O(\delta^{-2}\hat{k} + \mu)$. Thus, the event that there is a $t \leq \alpha N$ such that $V_t^{(\bar{k},\delta)} \neq V_t^{(\bar{k})}$ happens with probability $O(\delta^{-2}\hat{k} + \mu)$. □

We are now in a position to derive the desired propositions.

Proposition 6 *If $\bar{k} = O(\hat{k})$ and $0 < d = O(\hat{k})$ then*

$$P(|\tilde{U}(\bar{k}) + \tilde{V}(\bar{k})| > d) = O(d^{-2}\hat{k} + \mu).$$

Proof This follows directly from Lemmas 6 and 7 with $\delta = 2d$. □

Proposition 7 *If $\bar{k} = O(\hat{k})$, then, for $d > 0$,*

$$P(|\tilde{V}^{(\bar{k})}| > d) = O(\mu^{-1}d^{-2} + \mu \ln(N/\ell)).$$

Proof With $\delta = \mu^{-1}$, Lemmas 4 and 7 yield

$$P(|\tilde{V}^{(\bar{k})}| > d) = O(\delta d^{-2}(1 + \delta\mu)) + O(\delta^{-2}\hat{k} + \mu) = O(\mu^{-1}d^{-2} + \mu \ln(N/\ell)).$$

□

6 The proof of the main theorem

In this section we will use our previous achievements to finally prove Theorem 1. We will need the classical Berry–Esseen Theorem, which says how well the distribution of a sum of i.i.d. random variables is approximated by a normal distribution.

Theorem 2 (The Berry–Esseen Theorem) *Let W_1, W_2, \dots be independent and identically distributed random variables with $E(W_i) = 0$, $E(W_i^2) = \sigma^2 > 0$ and $E(|W_i|^3) = \tau < \infty$. Then, for any x and n ,*

$$|P(W_1 + \dots + W_n \leq x) - \Phi(x/(\sigma\sqrt{n}))| \leq \frac{\tau}{\sigma^3\sqrt{n}}.$$

At any point in time, let the *pseudolist* be the set of songs whose popularity is greater than or equal to \hat{k} .

Proposition 8 *The expected number L of songs that are on the pseudolist at a time t_0 but not at time $t_0 + \alpha N$ is $\sim \sqrt{\psi\alpha/\pi} \cdot \ell \cdot \sqrt{\mu \ln(N/\ell)}$.*

Proof Consider a song that has popularity $\bar{k} \geq \hat{k}$ at time t_0 , and define the random sequence $(K_t^{(\bar{k})})_{t=0}^\infty$ by letting $K_t^{(\bar{k})}$ be the popularity of the song at time $t_0 + t$. (Note that this definition is statistically equivalent to the definition of $K_t^{(\bar{k})}$ given in Sect. 5.) Let $Q_{\bar{k}} := P(K_{\alpha N}^{(\bar{k})} < \hat{k})$ be the probability that the song has left the pseudolist after αN time steps.

Introduce a variable v that tends to infinity but slowly enough so that $v\mu \ln(N/\ell) \rightarrow 0$ and $v = o(\sqrt{\ln(N/\ell)})$. (In fact, $v = o(\sqrt{\ln(N/\ell)})$ implies $v\mu \ln(N/\ell) \rightarrow 0$ via Assumption 2 but that does not matter.) We can divide L into three terms as follows.

$$L = \sum_{\bar{k}=\hat{k}}^N f_{\bar{k}} Q_{\bar{k}} = \overbrace{\sum_{\bar{k}=\hat{k}}^{\lfloor \hat{k}+v\sqrt{\hat{k}} \rfloor} f_{\bar{k}} Q_{\bar{k}}}^{L_1} + \overbrace{\sum_{\bar{k}=\lfloor \hat{k}+v\sqrt{\hat{k}} \rfloor+1}^{\lfloor \hat{k}/\psi \rfloor} f_{\bar{k}} Q_{\bar{k}}}^{L_2} + \overbrace{\sum_{\bar{k}=\lfloor \hat{k}/\psi \rfloor+1}^N f_{\bar{k}} Q_{\bar{k}}}^{L_3}.$$

First, we will deal with the term L_1 , so assume that $\bar{k} \in [\hat{k}, \hat{k} + v\sqrt{\hat{k}}]$. Let $d = \mu^{-1/2}(\ln(N/\ell))^{1/4}$ and define three events A_1, A_2 , and B as follows.

$$A_1 \iff \bar{k} + \tilde{U}^{(\bar{k})} < \hat{k} - d, \tag{43}$$

$$A_2 \iff \bar{k} + \tilde{U}^{(\bar{k})} < \hat{k} + d, \tag{44}$$

$$B \iff |\tilde{V}^{(\bar{k})}| > d. \tag{45}$$

We have the implications

$$A_1 \text{ and not } B \implies K_{\alpha N}^{(\bar{k})} < \hat{k} \implies A_2 \text{ or } B$$

and hence the inequalities

$$P(A_1) - P(B) \leq Q_{\bar{k}} \leq P(A_2) + P(B). \tag{46}$$

By Proposition 7, we have $P(B) = O(\mu^{-1}d^{-2} + \mu \ln(N/\ell))$ which is $o(v^{-1})$ with our choice of d and v .

Let

$$W_t := U_t^{(\bar{k})} - E(U_t^{(\bar{k})}) = \{(34)\} = U_t^{(\bar{k})} + \frac{\mu\bar{k}}{N}$$

and define

$$\begin{aligned} \sigma^2 &:= E(W_t^2) = \text{Var}(U_t^{(\bar{k})}) = \{(36)\} \\ &= \frac{\bar{k}}{N} \left(2 - \mu - 2(1 - \mu) \frac{\bar{k} - 1}{N - 1} \right) - \frac{\mu^2\bar{k}^2}{N^2} \sim \frac{2\bar{k}}{N} \end{aligned}$$

and

$$\begin{aligned} \tau &:= E(|W_t|^3) \\ &\leq E(|U_t^{(\bar{k})}|^3) + 3E((U_t^{(\bar{k})})^2)|E(U_t^{(\bar{k})})| + 3E(|U_t^{(\bar{k})}|)E(U_t^{(\bar{k})})^2 + |E(U_t^{(\bar{k})})|^3 \\ &\sim \frac{2\bar{k}}{N}. \end{aligned}$$

The Berry–Esseen Theorem now yields that, for any s ,

$$\left| P(\tilde{U}^{(\bar{k})} \leq s) - \Phi((s + \alpha\mu\bar{k})/(\sigma\sqrt{\alpha N})) \right| \lesssim \frac{1}{\sqrt{2\alpha\bar{k}}},$$

which implies, for $i = 1, 2$, that

$$\left| P(A_i) - \Phi((\hat{k} - \bar{k} + (-1)^i d + \alpha\mu\bar{k})/(\sigma\sqrt{\alpha N})) \right| = O(\bar{k}^{-1/2}). \tag{47}$$

Since $d = \mu^{-1/2}(\ln(N/\ell))^{1/4}$ and $\sigma^2 \sim 2\bar{k}/N = \Theta(\hat{k}/N)$, Corollary 1 yields that

$$\frac{d + \alpha\mu\bar{k}}{\sigma\sqrt{\alpha N}} \rightarrow 0.$$

Combining this with (46) and (47), we obtain

$$\left| Q_{\bar{k}} - \Phi\left((1 + o(1)) \frac{\hat{k} - \bar{k}}{\sqrt{2\alpha\hat{k}}} + o(1) \right) \right| = O(\hat{k}^{-1/2}) + P(B) = o(v^{-1}),$$

where the ordos converge uniformly over the interval $\bar{k} \in [\hat{k}, \hat{k} + v\sqrt{\hat{k}}]$. Summation over this interval yields

$$\left| \sum_{\bar{k}=\hat{k}}^{[\hat{k}+v\sqrt{\hat{k}}]} Q_{\bar{k}} - \sum_{\bar{k}=\hat{k}}^{[\hat{k}+v\sqrt{\hat{k}}]} \Phi\left((1 + o(1)) \frac{\hat{k} - \bar{k}}{\sqrt{2\alpha\hat{k}}} + o(1) \right) \right| = v\sqrt{\hat{k}} \cdot o(v^{-1}) = o(\sqrt{\hat{k}}).$$

By Proposition 4, $f_{\bar{k}} \sim \mu\ell$, so

$$\frac{L_1}{\mu\ell} \sim o(\sqrt{\hat{k}}) + \sum_{\bar{k}=\hat{k}}^{[\hat{k}+v\sqrt{\hat{k}}]} \Phi\left((1 + o(1)) \frac{\hat{k} - \bar{k}}{\sqrt{2\alpha\hat{k}}} + o(1) \right) \tag{48}$$

$$\sim o(\sqrt{\hat{k}}) + \int_{\bar{k}=\hat{k}}^{\hat{k}+v\sqrt{\hat{k}}} \Phi\left((1 + o(1)) \frac{\hat{k} - \bar{k}}{\sqrt{2\alpha\hat{k}}} + o(1) \right) d\bar{k} \tag{49}$$

$$= \left\{ x := \frac{\hat{k} - \bar{k}}{\sqrt{2\alpha\hat{k}}} \right\} \tag{50}$$

$$= o(\sqrt{\hat{k}}) + \sqrt{2\alpha\hat{k}} \int_{x=-v/\sqrt{2\alpha}}^0 \Phi((1 + o(1))x + o(1)) dx \tag{51}$$

$$\sim o(\sqrt{\hat{k}}) + \sqrt{2\alpha\hat{k}} \int_{x=-\infty}^0 \Phi(x) dx \tag{52}$$

$$\sim \frac{\sqrt{\alpha\hat{k}}}{\sqrt{\pi}}. \tag{53}$$

Next, we will deal with the term L_2 .

$$L_2 < f_{\hat{k}} \sum_{\bar{k}=[\hat{k}+v\sqrt{\hat{k}}]}^{[\hat{k}/\psi]} Q_{\bar{k}} = \{\text{Propositions 4 and 6}\} \tag{54}$$

$$= \mu\ell \sum_{\bar{k}=[\hat{k}+v\sqrt{\hat{k}}]}^{[\hat{k}/\psi]} O((\bar{k} - \hat{k})^{-2}\hat{k} + \mu) \tag{55}$$

$$= \mu\ell \cdot O\left(\int_{x=v\sqrt{\hat{k}}}^{(\psi^{-1}-1)\hat{k}} (x^{-2}\hat{k} + \mu) dx \right) \tag{56}$$

$$= \mu \ell \cdot O(\hat{k}(v^{-1}\hat{k}^{-1/2} - (\psi^{-1} - 1)^{-1}\hat{k}^{-1}) + \mu((\psi^{-1} - 1)\hat{k} - v\sqrt{\hat{k}})) \tag{57}$$

$$= \mu \ell \cdot (o(\sqrt{\hat{k}}) + \ln(N/\ell)) = \{\text{Corollary 1}\} \tag{58}$$

$$= \mu \ell \cdot o(\sqrt{\hat{k}}). \tag{59}$$

The last term, L_3 , is small simply because there are very few songs that are really popular:

$$\begin{aligned} \frac{L_3}{\ell\sqrt{\mu}} &\leq \frac{1}{\ell\sqrt{\mu}} \sum_{\bar{k}=\lfloor \hat{k}/\psi \rfloor + 1}^N f_{\bar{k}} \sim \{\text{Lemma 3}\} \\ &\sim \ell^{-1} \mu^{-3/2} f_{\lfloor \hat{k}/\psi \rfloor + 1} \sim \{\text{Lemma 2}\} \\ &\sim \frac{\psi N(1 - \mu)^{\hat{k}/\psi}}{\ell\sqrt{\mu} \cdot \hat{k}} \sim \{\text{Proposition 3}\} \\ &\sim \frac{\sqrt{\mu} \cdot N(1 - \mu)^{(1+o(1))\mu^{-1} \ln(N/\ell)}}{\ell \ln(N/\ell)} =: A. \end{aligned}$$

Taking the logarithm yields

$$\begin{aligned} \ln A &= -\frac{1}{2} \ln(\mu^{-1}) + \ln(N/\ell) + (1 + o(1))\mu^{-1} \ln(N/\ell) \ln(1 - \mu) - \ln \ln(N/\ell) \\ &= -\frac{1}{2} \ln(\mu^{-1}) + \ln(N/\ell) - (1 + o(1)) \ln(N/\ell) - \ln \ln(N/\ell) \\ &= \{\text{Assumption 2}\} \\ &= -\left(\frac{1}{2} + o(1)\right) \ln(N/\ell) + \ln(N/\ell) - (1 + o(1)) \ln(N/\ell) - \ln \ln(N/\ell) \\ &= -\left(\frac{1}{2} + o(1)\right) \ln(N/\ell) \rightarrow -\infty. \end{aligned}$$

Thus, $A \rightarrow 0$ and $L_3 = o(\ell\sqrt{\mu})$. □

In the end we are interested in the actual toplist rather than the pseudolist. The following proposition gives an upper bound for the difference between these lists.

Proposition 9 *Let S be the number of songs that are on exactly one of the toplist and the pseudolist. Then, the expected value $E(S) = O(\ell\sqrt{\mu} + \mu^{-1/2} \ln \mu^{-1})$.*

Proof Let $(\hat{k}, \hat{K}) := [\hat{k}, \hat{K}] \cup [\hat{K}, \hat{k}]$ denote the set of real numbers (inclusively) between \hat{k} and \hat{K} .

First, we overestimate S like this:

$$\begin{aligned}
 S \leq & \overbrace{\sum_{\substack{k \in \langle \hat{k}, \hat{K} \rangle \\ |k - \hat{k}| \leq \mu^{-1/2}}} X_k}^A + \overbrace{\sum_{\substack{k \in \langle \hat{k}, \hat{K} \rangle \\ \mu^{-1/2} < |k - \hat{k}| \leq \mu^{-3/4}}} X_k}^B \\
 & + \overbrace{1_{\hat{K} > \hat{k} + \mu^{-3/4}} \sum_{\hat{k} < k \leq N} X_k}^C + \overbrace{1_{\hat{K} < \hat{k} - \mu^{-3/4}} \ell}^D, \tag{60}
 \end{aligned}$$

where $1_{\text{“event”}}$ is an indicator variable for “event”. From Proposition 4 we recall that $E(X_k) = f_k \sim \mu \ell$ if $|k - \hat{k}| = o(\mu^{-1})$. Thus

$$E(A) \leq E\left(\sum_{|k - \hat{k}| < \mu^{-1/2}} X_k\right) = \sum_{|k - \hat{k}| < \mu^{-1/2}} f_k \sim \sum_{|k - \hat{k}| < \mu^{-1/2}} \mu \ell = O(\ell \sqrt{\mu}).$$

Now we will deal with the second sum, B , which can be written

$$B = \sum_{\mu^{-1/2} < |k - \hat{k}| \leq \mu^{-3/4}} X_k 1_{k \in \langle \hat{k}, \hat{K} \rangle}.$$

By the Cauchy–Schwartz inequality,

$$\begin{aligned}
 E(B) &= \sum_{\mu^{-1/2} < |k - \hat{k}| \leq \mu^{-3/4}} (f_k P(k \in \langle \hat{k}, \hat{K} \rangle) + \text{Cov}(X_k, 1_{k \in \langle \hat{k}, \hat{K} \rangle})) \\
 &\leq \sum_{\mu^{-1/2} < |k - \hat{k}| \leq \mu^{-3/4}} (f_k P(k \in \langle \hat{k}, \hat{K} \rangle) + \sqrt{\text{Var}(X_k) \text{Var}(1_{k \in \langle \hat{k}, \hat{K} \rangle})}).
 \end{aligned}$$

By Proposition 1, $\text{Var}(X_k) \leq f_k$, so we obtain

$$E(B) \leq \sum_{\mu^{-1/2} < |k - \hat{k}| \leq \mu^{-3/4}} (f_k P(k \in \langle \hat{k}, \hat{K} \rangle) + \sqrt{f_k P(k \in \langle \hat{k}, \hat{K} \rangle)}) \tag{61}$$

$$\sim \sum_{\mu^{-1/2} < |k - \hat{k}| \leq \mu^{-3/4}} (\mu \ell P(k \in \langle \hat{k}, \hat{K} \rangle) + \sqrt{\mu \ell P(k \in \langle \hat{k}, \hat{K} \rangle)}). \tag{62}$$

Proposition 5 with $\rho = |k - \hat{k}|$ yields

$$P(k \in \langle \hat{k}, \hat{K} \rangle) \leq P(|\hat{K} - \hat{k}| \geq |k - \hat{k}|) \lesssim 2|k - \hat{k}|^{-2} \mu^{-2} \ell^{-1},$$

and we obtain

$$\begin{aligned}
 E(B) &\lesssim 2 \sum_{\mu^{-1/2} < |k - \hat{k}| \leq \mu^{-3/4}} (\mu^{-1} |k - \hat{k}|^{-2} + \mu^{-1/2} |k - \hat{k}|) \\
 &\leq 4 \sum_{i=\lceil \mu^{-1/2} \rceil}^{\lfloor \mu^{-3/4} \rfloor} (\mu^{-1} i^{-2} + \mu^{-1/2} i^{-1}).
 \end{aligned}$$

In the last summand, the second term is larger than the first term over the summation interval, so

$$\begin{aligned}
 E(B) &\lesssim 8\mu^{-1/2} \sum_{i=\lceil \mu^{-1/2} \rceil}^{\lfloor \mu^{-3/4} \rfloor} i^{-1} \\
 &= O(\mu^{-1/2} \ln \mu^{-1}).
 \end{aligned}$$

By the Cauchy–Schwartz inequality,

$$\begin{aligned}
 E(C) &= E(1_{\hat{K} > \hat{k} + \mu^{-3/4}}) E\left(\sum_{\hat{k} < k \leq N} X_k\right) + \text{Cov}\left(1_{\hat{K} > \hat{k} + \mu^{-3/4}}, \sum_{\hat{k} < k \leq N} X_k\right) \\
 &\leq P(\hat{K} > \hat{k} + \mu^{-3/4}) E\left(\sum_{\hat{k} < k \leq N} X_k\right) \\
 &\quad + \sqrt{P(\hat{K} > \hat{k} + \mu^{-3/4}) \text{Var}\left(\sum_{\hat{k} < k \leq N} X_k\right)}.
 \end{aligned}$$

By definition of \hat{k} we have $E(\sum_{\hat{k} < k \leq N} X_k) < \ell$, so by Proposition 2, $\text{Var}(\sum_{\hat{k} < k \leq N} X_k) < \ell$. Proposition 5 with $\rho = \mu^{-3/4}$ yields $P(\hat{K} > \hat{k} + \mu^{-3/4}) = O(\mu^{-1/2} \ell^{-1})$. Thus,

$$E(C) = O(\mu^{-1/2} \ell^{-1} \ell + \sqrt{\mu^{-1/2} \ell^{-1} \ell}) = O(\mu^{-1/2}).$$

Finally, by Proposition 5, we get $E(D) = P(\hat{K} < \hat{k} - \mu^{-3/4}) \ell = O(\mu^{-1/2})$. \square

Now, at last, we are ready to prove Theorem 1.

Proof of Theorem 1 By Proposition 8 it suffices to show that the expected number $E(S)$ of songs that leave either the toplist or the pseudolist without leaving the other one, is much smaller than $\ell \sqrt{\mu \ln N} / \ell$. Proposition 9 tells us that $E(S) = O(\ell \sqrt{\mu} + \mu^{-1/2} \ln \mu^{-1})$ and the first term $\ell \sqrt{\mu}$ is clearly small enough. The second term is

$$\mu^{-1/2} \ln \mu^{-1} = \frac{\ln \mu^{-1}}{\ln(N/\ell)} \mu^{-1/2} \ln(N/\ell) \sim \{\text{Assumption 2}\} \quad (63)$$

$$\sim (1 - \psi) \mu^{-1/2} \ln(N/\ell) < \mu^{-1/2} \sqrt{\ln(N/\ell)} \sqrt{\ln(N/\ell)} \quad (64)$$

$$\ll \{\text{Assumption 3}\} \ll \mu^{-1/2} \mu \ell \sqrt{\ln(N/\ell)}. \quad (65)$$

□

7 Discussion

Bentley et al. [3] conjectured a simple expression for the turnover rate of popularity tolists in a random copying model with nonoverlapping generations (the infinite alleles Wright–Fisher model). In this paper we instead studied the overlapping generations version, known as the infinite alleles Moran model. We first showed by simulations that the toplist turnover rate seems to behave in the same way for the two models (for the appropriate regime of short lists compared to the population size). We then proved an asymptotic formula for the turnover rate, which modifies the conjectured formula by a factor $\sqrt{\ln(N/\ell)}$. In other words, the turnover rate is not perfectly independent of the population size N , but the dependence will not be noticeable in data unless one considers truly huge variations of N .

It is interesting that the two models behave so similarly with respect to toplist turnover. It is worth investigating how robust this behavior is to other reasonable changes of the model. For instance, there may be various forms of biases to the random copying, as discussed by Boyd and Richerson [4]. For instance, some pop songs may actually be better than others in some sense that makes them more likely to be voted for. Boyd and Richerson also discuss frequency-dependent biases, that would make already popular songs more (or less) likely to be voted for.

Another type of change is to let the population be increasing rather than fixed. For instance, if we remove the death step from the IAM model, it becomes equivalent to economist Herbert Simon's famous model of urban growth [7], for which toplist turnover would certainly be a relevant aspect to study. The book of Andrews and Eriksson [2] discusses a couple of other important random growth processes on Young diagrams for which the same question could be asked.

The toplist turnover problem seems to be novel. We envisage that a broader mathematical investigation of it may be fruitful.

References

1. Andrews, G.E., Berndt, B.C.: Your hit parade: The top ten most fascinating formulas in Ramanujan's lost notebook. *Not. Am. Math. Soc.* **55**, 18–30 (2008)
2. Andrews, G.E., Eriksson, K.: *Integer Partitions*. Cambridge University Press, Cambridge (2004)
3. Bentley, R.A., Lipo, C.P., Herzog, H.A., Hahn, M.W.: Regular rates of popular culture change reflect random copying. *Evol. Human Behav.* **28**, 151–158 (2007)
4. Boyd, R., Richerson, P.J.: *Culture and the Evolutionary Process*. University of Chicago Press, Chicago (1985)

5. Eriksson, K., Sjöstrand, J.: Limiting shapes of birth-and-death processes on Young diagrams. *Adv. Appl. Math.* (in press)
6. Ewens, W.J.: *Mathematical Population Genetics*, 2nd edn. Springer, New York (2004)
7. Simon, H.: On a class of skew distribution functions. *Biometrika* **42**, 425–440 (1955)
8. Strimling, P., Sjöstrand, J., Eriksson, K., Enquist, M.: Accumulation of independent cultural traits. *Theor. Popul. Biol.* **76**, 77–83 (2009)