

Population Dynamic Models Leading to Logarithmic and Yule Distribution

János Izsák¹, László Szeidl²

¹Department of Systematical Zoology and Ecology, Eötvös Loránd University, Pázmány Péter sétány 1/C, H-1117 Budapest, Hungary

E-mail: ijanos@caesar.elte.hu

²Institute of Applied Informatics, John von Neumann Faculty of Informatics, Óbuda University, Bécsi u. 96/B. H-1034 Budapest, Hungary

E-mail: szeidl@uni-obuda.hu

Abstract: A significant field of species abundance distribution (SAD) has a population dynamical character, in which it is supposed that the stochastic speciation process and the evolution of different species are determined by the same linear birth and death process. The distributions of the number of individuals after the speciation tend to a discrete limit distribution depending on some condition if the observation time increases. In the earlier publications, in general, the speciation process was supposed to be a homogeneous Poisson process. In a more realistic case, if the speciation process is inhomogeneous Poisson, the investigation of the model is obviously more difficult. In this paper we deal with the models, in which the birth and death intensities are identical, the speciation rate is bounded, locally integrable and has asymptotically power type behaviour. Limit parameters for these models, depending on the speciation rate, are proportional to a logarithmic or (exactly or asymptotically) Yule distribution. In connection with the sample statistics some results are derived in general and also in special cases (logarithmic and Yule distribution), which are related to the random choice of a species or an individual from the whole population of the system.

Keywords: population dynamic model; species abundance distribution; Kendall process; Poisson process; logarithmic distribution; Yule distribution

1 Introduction

A frequently cited field of species abundance models possesses population dynamical background. In these models continuous abundances are mostly assumed (Engen

and Lande, 1996a, 1996b). Generally, it is supposed that the process which describes the entering time points of the new species in the system is Poisson process (Karlin és McGregor, 1967). In this paper we consider models in which the species abundance can take discrete values $0, 1, 2, \dots$, the evolution of the species entering the system is determined by a linear birth and death model (Kendall (1948a, 1948b) and as an essential enlargement of the population dynamical models, the speciation processes are assumed from a class of inhomogeneous Poisson processes. For the description of the model parameters, the Yule distribution plays an important role instead of logarithmic distribution.

It is worth noting that in case when the speciation rate is varying in time, i.e. the speciation process in the model is inhomogeneous Poisson, it is more difficult to reach concrete results. We mention here the results of Branson (1991, 2000), in which the models lead to logarithmic distribution under special inhomogeneity condition. In this paper we deal with a class of models which lead to distributions exactly or asymptotically proportional to the Yule distribution.

Note that the Yule distribution with parameter $\rho (> 0)$ is determined as $p_k = \rho \Gamma(\rho + 1) \frac{\Gamma(k)}{\Gamma(k + \rho + 1)}$, $k = 1, 2, \dots$, for which the asymptotic relation $p_k = \rho \Gamma(\rho + 1) k^{-\rho - 1} (1 + o(1))$, $k \rightarrow \infty$ holds and it can be interpreted as a generalization of power type (Pareto type) distribution for a discrete case (see Simon (1955), Newman (2006)).

Let us consider a system of many species on the time interval $(t_0 - T, t_0]$, where $t_0 \leq 0$, $T > 0$ and the system is empty at the initial time $t_0 - T$, i.e. the system does not contain any species. After passing T time we investigate the system at the observation time t_0 . The time points, when the species enter in the system, are determined by the random jumping points of a homogeneous or inhomogeneous Poisson process Π , having intensity function $\lambda(t)$, $t \leq t_0$, which is defined on the half line $(-\infty, t_0]$ (see 4.5.§., Kingman (1993)). The process Π defines a right continuous Poisson process $N_T(t)$, $t_0 - T \leq t \leq t_0$ for each $T > 0$ on the time interval $(t_0 - T, t_0]$ satisfying the condition $N_T(t_0 - t) = 0$. We mention that the process $N_T(t)$ can be given by construction (see p. 50., Kingman (1993), p. 62., Lakatos et al. (2013)).

It is clear that the Poisson process $N_T(t)$, $t_0 - T \leq t \leq t_0$ has rate function which equals $\lambda(t)$ on the interval $t_0 - T \leq t \leq t_0$. The rate function (of formation of a new species) $\lambda(t)$ of the speciation process does not depend on the species entering the system, but it can depend on time t . Then for any pairwise disjoint intervals $(x_i, y_i] \subset (t_0 - T, t_0]$, $i = 1, 2, \dots$ the increments $N_T(y_i) - N_T(x_i)$ are independent random variables with Poisson distribution of parameter $E(N_T(y_i) - N_T(x_i)) = \int_{x_i}^{y_i} \lambda(s) ds$.

Note that if we investigate the abundance distribution for the case of homogeneous (i.e. $\lambda(t) \equiv \lambda_0$) speciation process then we have the same distribution for any observation time t_0 as $T \rightarrow \infty$, in contrast with the inhomogeneous cases, when the limit depends on the observation time t_0 . Partly, this means that if there exists the limit abundance distribution in homogeneous cases as $T \rightarrow \infty$ then the limit is identical with the equilibrium (stationary) distribution, while in cases of inhomogeneous speciation process this property is no longer valid.

Assume that the number of individuals of a species entering the system equals 1. Moreover, the random fluctuation of the population size of a species does not depend on others and it is determined by a continuous-time Markov chain for all species with the same transition probabilities $P_{1,k}(s)$, $s \geq 0$, $k = 0, 1, \dots$. The state 0 (i.e. the extinction of a species) means the absorption state. Then the number of species $S_{k,T}$, $k = 1, 2, \dots$ having exactly k , $k = 1, 2, \dots$ living individuals at the observation time t_0 are independent random variables with Poisson distribution of parameters $\mu_{k,T}$, $k = 1, 2, \dots$ which can be given in the following general form (Karlin and McGregor, 1967)

$$\mu_{k,T} = \int_{t_0-T}^{t_0} P_{1,k}(t_0-t)\lambda(t)dt = \int_0^T P_{1,k}(t)\lambda(-t+t_0)dt, \quad k = 1, 2, \dots \quad (1)$$

This formula plays an important role in the computation of the parameters $\mu_{k,T}$. In accordance with the Kendall population dynamical model, after a species enters the system, the random fluctuation of the population size of a species is determined by a linear birth and death model (Kendall, 1948a, 1948b), where the birth and death rates na and nc , respectively, depend on the actual population size n of the species and a and c ($a \leq c$) are positive constants.

It is known (see Karlin and McGregor (1967)) that if the speciation process $N_T(t)$ is homogeneous Poisson with intensity rate λ , then the random variables $S_{k,T}$, $k = 1, 2, \dots$ (the number of species $S_{k,T}$, $k = 1, 2, \dots$ having exactly k , $k = 1, 2, \dots$ living individuals) are independent and have Poisson distribution with parameters $\mu_{k,T}$ (see also Engen and Lande (1996a, 1996b), Watterson (1974), Lange (2010), Bowler and Kelly (2012)), where

$$\mu_{k,T} = \frac{\lambda}{a} \frac{1}{k} \rho^k \left(\frac{1 - e^{-(c-a)T}}{1 - \rho e^{-(c-a)T}} \right)^k \rightarrow \mu_k = \frac{\lambda}{a} \frac{1}{k} \rho^k, \quad T \rightarrow \infty, \quad \text{if } \rho = a/c < 1 \quad (2)$$

and

$$\mu_{k,T} = \frac{\lambda}{a} \frac{1}{k} \left(\frac{aT}{1+aT} \right)^k \rightarrow \mu_k = \frac{\lambda}{a} \frac{1}{k}, \quad T \rightarrow \infty, \quad \text{if } a = c. \quad (3)$$

From the formulas (2) and (3) it follows for the case $a < c$ ($\rho < 1$) that the sequence $\mu_{k,T}$, $k = 1, 2, \dots$ (i.e. the expected values of the number of species having exactly k individuals) is proportional to the logarithmic distribution with parameter $\rho \frac{1 - e^{-(c-a)T}}{1 - \rho e^{-(c-a)T}}$ (in the limit as $T \rightarrow \infty$ with parameter ρ). This distribution does not depend on the observation time t_0 if the speciation process is homogeneous, i.e. $\lambda(t) \equiv \lambda_0$.

In case $a = c$ the sequence $\mu_{k,T}$, $k = 1, 2, \dots$ is proportional to a logarithmic distribution only if $T < \infty$. In that case the parameter of the logarithmic distribution equals

$\frac{aT}{1+aT}$. The parameters $\mu_{k,T}$ have the limit

$$\mu_k = \lim_{T \rightarrow \infty} \mu_{k,T} = \frac{\lambda_0}{a} \frac{1}{k}, \quad k = 1, 2, \dots,$$

however, they will be no longer proportional to a probability distribution because $\sum_{k=1}^{\infty} \frac{1}{k} = \infty$. From this it follows that if $T \rightarrow \infty$, then the expected value of the number of species having minimum one individual at time t_0 , tends to ∞ , at the same time the expected value of the number of species with exactly k individuals tends to value $\frac{\lambda_0}{a} \frac{1}{k}$. Thus the number of species with exactly k individuals has Poisson limit distribution with parameter $\frac{\lambda_0}{a} \frac{1}{k}, k = 1, 2, \dots$

We note that in the remaining case under the condition $a > c$ for all $k = 1, 2, \dots$ $\lim_{T \rightarrow \infty} ES_{k,T} = \lim_{T \rightarrow \infty} \mu_{k,T} = \infty$ is true.

2 Results

In the present section of the paper we study two problems, as follows.

1. We consider the birth and death process under the condition that the birth and death rates are equal ($a = c$), however, the rate $\lambda(t)$ of the Poisson speciation process $N(t)$ is inhomogeneous. The problem is to give exact and asymptotic formulas for the behaviour of the parameters $\mu_{k,T}$ as $T \rightarrow \infty$ under the condition

$$\lambda(t) = \frac{\lambda_0}{(1 + \alpha|t|)^\beta}, \quad -\infty < t \leq 0 \tag{4}$$

or in more general setting, if $\lambda(t)$ satisfies the asymptotic condition

$$\frac{(1 + \alpha|t|)^\beta}{\lambda_0} \lambda(t) \rightarrow 1, \quad t \rightarrow -\infty, \quad \lambda_0, \alpha, \beta > 0, \quad -\infty < t \leq t_0 \leq 0, \tag{5}$$

where λ_0, α and β are arbitrary positive numbers. This model generalizes the above described models.

2. In connection to this model, we consider a random choice problem for Poissonian distributed abundances at observation time t_0 . In this model, we investigate a species randomly chosen from the population or an individual from the whole population with which probability belongs to a species with k ($k \geq 1$) individuals.

2.1 Exact and asymptotic results for the parameters $\mu_k = \lim_{T \rightarrow \infty} \mu_{k,T}$ when the speciation rate $\lambda(t)$ satisfies the conditions (4) and (5)

In case of inhomogeneous Poisson speciation process, the consideration at time t_0 of the parameters μ_k will be more difficult comparing to a homogeneous case, be-

cause the parameters $\mu_k = \int_{-\infty}^{t_0} P_{1k}(t_0 - t)\lambda(t)dt$, $k = 1, 2, \dots$ may depend not only on speciation rate $\lambda(t)$ but also on the observation time t_0 .

In this section we assume that the condition (4) or (5) holds, instead of the homogeneity of the speciation rate ($\lambda(t) \equiv \lambda_0$), which makes possible a more general framework for the modelling of the population dynamics. Here the observation time $t_0 \leq 0$ can be arbitrarily chosen. Note that under the condition (4) $\lambda(t)$ is a monotonically increasing function which realizes monotonically increasing speciation rate. The fact that the speciation rate can be increasing, from a biological point of view, is referred in the paper of Rolland et al. (2014). In special cases we give exact formulas for the parameters μ_k , $k = 1, 2, \dots$, and at the same time the asymptotic formulas will be valid for the class of bounded rate functions $\lambda(t)$ satisfying the more general condition (5), instead of (4).

In accordance with the model stated above, the dynamics (in time) of the number of individuals of a species is described by a linear birth and death process (Kendall process) for which the rate of birth and death are na and nc , respectively, depending on the population size n and on the given constants $a, c > 0$. The initial population size of a species is 1 and the state 0 is an absorbing one. The birth and death process is a continuous-time Markov chain, which determines the random fluctuation of the population size in time after speciation.

Denote the population size of species by X_t , $t \geq 0$, $X_0 = 1$, where t means the passing time after speciation and let $P_{1k}(t) = P(X_t = k | X_0 = 1)$, $k = 0, 1, \dots$ be the transition probability function of the process. Since the initial state of the process is 1, thus $P_{11}(0) = 1$ and $P_{1k}(0) = 0$, $k \neq 1$.

The generating function of the time-dependent transition probabilities $P_{1k}(t)$, $k = 0, 1, \dots$ of the Markov chain X_t , $t \geq 0$ can be determined by the Kolmogorov forward differential equations, from which the transition probabilities $P_{1k}(t)$ can be given in an explicit form (Kendall, 1948a):

$$P_{1,0}(t) = \frac{at}{1+at}, \quad P_{1,k}(t) = \frac{(at)^{k-1}}{(1+at)^{k+1}}, \quad k = 1, 2, \dots \quad (6)$$

Theorem 1. If the birth and death intensities are equal ($a = c$) and the intensity function of the speciation process satisfies the condition (5), then

a) independently of the value t_0 the following asymptotic relation holds

$$\mu_k = \frac{\lambda_0}{a} \left(\frac{a}{\alpha}\right)^\beta \beta \Gamma(\beta + 1) \frac{1}{k^{\beta+1}} (1 + (1)), \quad k \rightarrow \infty. \quad (7)$$

This means that for sufficiently large k , the elements of the sequence μ_k of expected values of the numbers of the species with k members are asymptotically proportional to the elements of a Yule distribution with parameter β .

b) Under the condition (4) an exact formula holds for the sequence μ_k , $k = 1, 2, \dots$ if $a = \alpha > 0$, $\beta > 0$ and $t_0 = 0$ is the time of the observation. In this case the sequence μ_k , $k = 1, 2, \dots$ can be given with the help of the Yule distribution of parameter β

multiplying by the constant $\frac{\lambda_0}{a\beta}$ as follows:

$$\mu_k = \frac{\lambda_0}{a\beta} \beta \frac{\Gamma(k)\Gamma(\beta + 1)}{\Gamma(k + \beta + 1)}, \quad k = 1, 2, \dots \tag{8}$$

In the special case, for $\beta = 1$ the equation $\mu_k = \frac{\lambda_0}{a} \frac{1}{k(k+1)}$, $k = 1, 2, \dots$, holds and for $\beta = 2$ the equation $\mu_k = \frac{\lambda_0}{a} \frac{2}{k(k+1)(k+2)}$, $k = 1, 2, \dots$ is true.

Proof. For simplicity, define $\lambda(t) = \lambda(-t)$, $t > 0$. If the birth and death rates are equal, i.e. $a = c$, then the transition probabilities $P_{1k}(t)$ satisfy the relations (6), therefore by the formula (1) the numbers of species with $k \geq 1$ members at the observation time t_0 ($t_0 \leq 0$) are independent and have Poisson distribution with parameters (expected values) as follows

$$\begin{aligned} \mu_k &= \int_{-\infty}^{t_0} P_{1k}(t_0 - t)\lambda(t)dt = \int_{-\infty}^0 P_{1k}(-t)\lambda(t_0 + t)dt = \int_0^{\infty} P_{1k}(t)\lambda(t_0 - t)dt = \\ &= \int_0^{\infty} \frac{(at)^{k-1}}{(1 + at)^{k+1}} \lambda(t - t_0)dt, \quad k = 1, 2, \dots \end{aligned} \tag{9}$$

These integrals are finite because the integrands are bounded, moreover, from the condition (5) $\frac{(1+\alpha|t|)^\beta}{\lambda_0} \lambda(t) \rightarrow 1$, $t \rightarrow \infty$ follows, then by (6) we have

$$P_{1,k}(t)\lambda(t) = \frac{\lambda_0}{\alpha^\beta} t^{-\beta-2}(1 + o(1)), \quad t \rightarrow \infty,$$

which means that $(P_{1,k}(t)\lambda(t))^{-1} \frac{\lambda_0}{\alpha^\beta} t^{-\beta-2} \rightarrow 1$, $t \rightarrow \infty$. The integral in (9) can be given in the form

$$\mu_k = \frac{\lambda_0}{a} \int_0^{\infty} f_k(t)g(t)dt, \quad k \geq 1, \tag{10}$$

where

$$f_k(t) = \frac{t^{k-1}}{(1 + t)^{k+1+\beta}}, \quad g(t) = \frac{1}{\lambda_0} (1 + t)^\beta \lambda(t/a + |t_0|).$$

It is clear that from the condition (5) it follows that the function $g(t)$ satisfies the asymptotic relation

$$g(t) = \frac{(1 + t)^\beta}{[1 + \alpha(t/a + |t_0|)]^\beta} \frac{1}{\lambda_0} [1 + \alpha(t/a + |t_0|)]^\beta \lambda(t/a + |t_0|) \rightarrow \left(\frac{a}{\alpha}\right)^\beta, \quad t \rightarrow \infty. \tag{11}$$

Let us consider the asymptotic behaviour of the parameters μ_k as $k \rightarrow \infty$. Firstly, we prove that the following convergence is true

$$\left(\int_0^{\infty} f_k(t)dt\right)^{-1} \mu_k \rightarrow \frac{\lambda_0}{a} \left(\frac{a}{\alpha}\right)^\beta, \quad k \rightarrow \infty. \tag{12}$$

Since the integral $\int_0^\infty f_k(t)dt$, $k = 1, 2, \dots$ in formula (12) can be determined by formula 2.2.4.24., p. 298., Prudnikov et al. (1986) and it equals the Yule distribution of parameter β as follows

$$\int_0^\infty f_k(t)dt = \int_0^\infty \frac{t^{k-1}}{(1+t)^{k+1+\beta}} dt = \frac{\Gamma(k)\Gamma(\beta+1)}{\Gamma(k+\beta+1)}, k \geq 1, \quad (13)$$

therefore if we prove the relations (12) and (13) we immediately have the asymptotic relation (7) of the Theorem 1.

It is known that the gamma function has the following asymptotic property (see p. 257, Davis, 1972): for any fixed real numbers u, v

$$\frac{\Gamma(x+u)}{\Gamma(x+v)} = x^{u-v}(1+o(1)), x \rightarrow \infty, \quad (14)$$

consequently, by (13) and (14) we have

$$\int_0^\infty f_k(t)dt = \Gamma(\beta+1) \frac{1}{k^{\beta+1}}(1+o(1)), k \rightarrow \infty. \quad (15)$$

Now, we verify the relation (12). For arbitrary positive numbers γ, A and for any $0 \leq t \leq A$

$$\begin{aligned} k^\gamma \left(\frac{t}{1+t} \right)^k &\leq k^\gamma \left(\frac{A}{1+A} \right)^k = \exp \left\{ k \log \frac{A}{1+A} + \gamma \log k \right\} = \\ &= \exp \left\{ -k \left[\log \left(1 + \frac{1}{A} \right) - \frac{\gamma}{k} \log k \right] \right\} \rightarrow 0, k \rightarrow \infty \end{aligned}$$

holds. It is obvious that

$$\int_0^A f_k(t)dt < \left(\frac{A}{1+A} \right)^{k-1} \int_0^A \frac{1}{(1+t)^{\beta+2}} dt < \left(\frac{A}{1+A} \right)^{k-1} \rightarrow 0, k \rightarrow \infty.$$

Since $\lambda(t)$ and consequently $g(t)$ are bounded functions, then for $\gamma = \beta + 1$ we have

$$k^{\beta+1} \int_0^A f_k(t)g(t)dt < \max_{0 \leq t \leq A} g(t) \cdot \left(\frac{A}{1+A} \right)^{k-1} \rightarrow 0, k \rightarrow \infty \quad (16)$$

and

$$k^{\beta+1} \int_0^A f_k(t)dt < k^{\beta+1} \left(\frac{A}{1+A} \right)^{k-1} \rightarrow 0, k \rightarrow \infty. \quad (17)$$

By virtue of the asymptotic relation (11) the convergence $g(t) \rightarrow \left(\frac{a}{\alpha} \right)^\beta$, $t \rightarrow \infty$ is true, therefore for arbitrarily chosen $\varepsilon > 0$ there exists a constant A_ε such that

$$\left| g(t) - \left(\frac{a}{\alpha} \right)^\beta \right| < \varepsilon, t \geq A_\varepsilon. \quad (18)$$

From this it follows that

$$\left| \int_{A_\varepsilon}^\infty f_k(t) \left(\frac{a}{\alpha}\right)^\beta dt - \int_{A_\varepsilon}^\infty f_k(t)g(t)dt \right| \leq \varepsilon \int_{A_\varepsilon}^\infty f_k(t)dt. \tag{19}$$

In summary, on the basis of the relations (15), (16) and (17) from (19) it is clear that for every $\varepsilon > 0$ it holds

$$\limsup_{k \rightarrow \infty} \left| \left(\int_0^\infty f_k(t)dt \right)^{-1} \int_0^\infty f_k(t)g(t)dt - \left(\frac{a}{\alpha}\right)^\beta \right| < \varepsilon,$$

thus

$$\mu_k = \frac{\lambda_0}{a} \int_0^\infty f_k(t)g(t)dt = \frac{\lambda_0}{a} \left(\frac{a}{\alpha}\right)^\beta \Gamma(\beta + 1) \frac{1}{k^{\beta+1}} (1 + o(1)), \quad k \rightarrow \infty.$$

The result of the second part b) of the Theorem 1 is obtained directly from the formulas (9) and (13):

$$\mu_k = \frac{\lambda_0}{a} \int_0^\infty \frac{t^{k-1}}{(1+t)^{k+1+\beta}} dt = \frac{\lambda_0}{a} \frac{\Gamma(k)\Gamma(\beta + 1)}{\Gamma(k + \beta + 1)}.$$

□

2.2 Theorems on the random choice of a species or an individual from the whole population related to the model considered above

Consider a population of various species. Assume in general that the number of species of the population is not necessarily bounded. Denote the number of species consisting of exactly k individuals by S_k , $k = 1, 2, \dots$ and suppose that the random variables S_k are independent, having Poisson distribution with parameters μ_k , $k = 1, 2, \dots$ and the condition $\mu = \sum_{k=1}^\infty \mu_k < \infty$ holds. For example, the random variables may be S_k , the number of species having k individuals at the time t_0 (see the model described earlier). Define the events A_k and B_k as follows

$A_k = \{\text{randomly chosen species from the population of species consists of } k$
individuals}\},

$B_k = \{\text{randomly chosen individual from the population of individuals belongs to}$
a species consisting of exactly } k \text{ individuals}\}.

Let us consider the probabilities $P(A_k)$, and $P(B_k)$, $k = 1, 2, \dots$ of the events A_k , and B_k , respectively. Denote $S_k = \sum_{i \neq k} S_i$ and $R_k = \frac{1}{k} \sum_{i \neq k} i S_i$, $k = 1, 2, \dots$. Let \mathcal{R}_k be the set of all possible values of the random variables $kR_k = \sum_{i \neq k} i S_i$, that is, for $k = 1, 2, \dots$

$\mathcal{R}_k = \{ \sum_{i \neq k} im_i : m_i \text{ are arbitrary natural numbers and the sum } \sum_{i \neq k} im_i \text{ is finite} \}$.

The random choice of a species or an individual from the population considered above means that for all $n \geq 0$, $m \geq 0$, $n + m > 0$ and $r \in \mathcal{R}_k$ the following relations hold

$$P(A_k | S_k = n, \bar{S}_k = m) = \frac{n}{n+m}, \quad P(B_k | S_k = n, R_k = \frac{1}{k}r) = \frac{kn}{kn+r}.$$

Using the formula of total probability we get

$$\begin{aligned} P(A_k) &= \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} P(A_k | S_k = n, \bar{S}_k = m) P(S_k = n, \bar{S}_k = m) = \\ &= \sum_{n=1}^{\infty} \sum_{m=0}^{\infty} \frac{n}{n+m} P(S_k = n, \bar{S}_k = m) = E \frac{S_k}{S_k + \bar{S}_k}. \end{aligned} \quad (20)$$

Taking into consideration that the random variables S_k and \bar{S}_k are independent and have Poisson distribution with parameters μ_k and $\mu - \mu_k$ respectively, using the relation (20) it is easy to determine the well-known general formula (21) for the probability $P(A_k)$

$$P(A_k) = \frac{\mu_k}{\mu}, \quad k = 1, 2, \dots \quad (21)$$

The computation of the probability $P(B_k)$, $k = 1, 2, \dots$ is more difficult and leads to an interesting formula determined by the parameters μ_k , μ and the generating function (z -transform) $G(z)$ of the sequence μ_k , $k = 1, 2, \dots$. This formula makes the further consideration of the probability $P(B_k)$ as $k \rightarrow \infty$ possible.

The number of different species possessing the population is $S_k + \bar{S}_k = \sum_{i=1}^{\infty} S_i$ and the number of individuals in the population equals $kS_k + kR_k = \sum_{i=1}^{\infty} iS_i$. Using the formula of the total probability we have

$$\begin{aligned} P(B_k) &= \sum_{n=0}^{\infty} \sum_{r \in \mathcal{R}_k} P(B_k | S_k = n, kR_k = r) P(S_k = n, kR_k = r) = \\ &= \sum_{n=1}^{\infty} \sum_{r \in \mathcal{R}_k} \frac{kn}{kn+r} P(S_k = n, kR_k = r) = E \frac{S_k}{S_k + R_k}. \end{aligned} \quad (22)$$

It will be noted that $\sum_{n=1}^{\infty} P(B_k) = 1$, because

$$\sum_{n=1}^{\infty} E \frac{S_k}{S_k + R_k} = E \sum_{n=1}^{\infty} \frac{kS_k}{kS_k + kR_k} = E \frac{kS_k + kR_k}{kS_k + kR_k} = 1.$$

Let us define the generating function of $G(z)$ of the sequence μ_k , $k = 1, 2, \dots$ as follows

$$G(z) = \sum_{k=1}^{\infty} \mu_k z^k, \quad |z| \leq 1.$$

Theorem 2. If $S_i, i = 1, 2, \dots$ denote the number of species containing i individuals and the random variables S_1, S_2, \dots are independent and they have Poisson distribution function with parameters μ_1, μ_2, \dots , then the following relation holds

$$P(B_k) = \mu_k \int_0^1 \exp \left\{ - \sum_{i=1}^{\infty} \mu_i (1 - x^{i/k}) \right\} dx = \mu_k \int_0^1 \exp \left\{ -\mu + G(x^{1/k}) \right\} dx \quad (23)$$

and the probabilities $P(B_k)$ satisfy the asymptotic relation

$$P(B_k) = \mu_k(1 + o(1)), \quad k \rightarrow \infty. \quad (24)$$

Proof. Since the random variables S_k and R_k are independent, then using the formula (22) the probability $P(B_k)$ can be given in the form

$$P(B_k) = E \frac{S_k}{S_k + R_k} = E \left(E \left(\frac{S_k}{S_k + R_k} \mid R_k \right) \right) = E \left(\sum_{n=1}^{\infty} \frac{n}{n + R_k} \frac{\mu_k^n}{n!} e^{-\mu_k} \right).$$

It is clear that $P(B_k) = 0$, when $\mu_k = 0$ and for $\mu_k > 0$

$$\frac{\mu_k^n}{n + R_k} = \mu_k^{-R_k} \int_0^{\mu_k} x^{R_k+n-1} dx.$$

The order of summation and integration, as well as the order of integration and expectation can be changed in the following relation, thus we have

$$\begin{aligned} P(B_k) &= e^{-\mu_k} E \left(\sum_{n=1}^{\infty} \frac{1}{(n-1)!} \mu_k^{-R_k} \int_0^{\mu_k} x^{R_k+n-1} dx \right) = \\ &= e^{-\mu_k} E \left(\int_0^{\mu_k} \mu_k^{-R_k} x^{R_k} \sum_{n=0}^{\infty} \frac{x^n}{n!} dx \right) = e^{-\mu_k} E \left(\int_0^{\mu_k} \left(\frac{x}{\mu_k} \right)^{R_k} e^x dx \right) = \\ &= e^{-\mu_k} \mu_k E \left(\int_0^1 x^{R_k} e^{\mu_k x} dx \right) = e^{-\mu_k} \mu_k \int_0^1 E(x^{R_k}) e^{\mu_k x} dx. \end{aligned} \quad (25)$$

The expected value $E x^{R_k}$ equals the generating function of random variable $R_k = \sum_{i \neq k} \frac{i}{k} S_i$ in the place $x^{1/k}$, which is easy to compute. Since the random variables S_i ($i = 1, 2, \dots$) are independent and they have Poisson distribution with parameters μ_i , $i = 1, 2, \dots$, moreover, the generating function of random variable S_i has the form

$$E x^{S_i} = e^{\mu_i(x-1)}, \quad 0 < x \leq 1,$$

then

$$\begin{aligned} E x^{R_k} &= E(x^{1/k})^{\sum_{i \neq k} i S_i} = E \prod_{i \neq k} \left(x^{i/k} \right)^{S_i} = \exp \left\{ \sum_{i \neq k} \mu_i (x^{i/k} - 1) \right\} \\ &= \exp \left\{ -(\mu - \mu_k) + \sum_{i \neq k} \mu_i x^{i/k} \right\}. \end{aligned}$$

From the formula (25) we get

$$\begin{aligned} P(B_k) &= e^{-\mu_k} \mu_k \int_0^1 \exp \left\{ -(\mu - \mu_k) + \sum_{i \neq k} \mu_i x^{i/k} \right\} e^{\mu_k x} dx = \\ &= \mu_k \int_0^1 \exp \left\{ -\mu + \sum_{i=1}^{\infty} \mu_i x^{i/k} \right\} dx = \mu_k \int_0^1 \exp \left\{ -\mu + G(x^{1/k}) \right\} dx, \end{aligned}$$

which is the statement (23) of the Theorem.

We now prove that the asymptotic relation (24) holds. Using the formula (23) it is enough to verify that the following convergence holds

$$\int_0^1 \exp \left\{ -\mu + G(x^{1/k}) \right\} dx \rightarrow 1, \text{ if } k \rightarrow \infty. \quad (26)$$

On the one hand, the generating function $G(x)$ is continuous, monotonically increasing on the interval $[0, 1]$ and has the limit value μ from left in the point 1, then $0 \leq \mu - G(x^{1/k}) \leq \mu - G(\varepsilon^{1/k})$, $0 \leq \varepsilon \leq x \leq 1$. On the other hand, for every fixed constant ε , $0 < \varepsilon < 1$ the convergence $\varepsilon^{1/k} \rightarrow 1$ holds as $k \rightarrow \infty$, then $G(\varepsilon^{1/k}) \rightarrow \mu$, $k \rightarrow \infty$ and

$$\begin{aligned} 1 &\geq \int_0^1 \exp \left\{ -\mu + G(x^{1/k}) \right\} dx = \\ &= \int_0^{\varepsilon} \exp \left\{ -\mu + G(x^{1/k}) \right\} dx + \int_{\varepsilon}^1 \exp \left\{ -\mu + G(\varepsilon^{1/k}) \right\} dx \geq \\ &\geq \varepsilon e^{-\mu} + (1 - \varepsilon) \exp \left\{ -\mu + G(\varepsilon^{1/k}) \right\}. \end{aligned}$$

Since the constant ε , $0 < \varepsilon < 1$ can be arbitrarily chosen and

$$\varepsilon e^{-\mu} + (1 - \varepsilon) \exp \left\{ -\mu + G(\varepsilon^{1/k}) \right\} \rightarrow \varepsilon e^{-\mu} + (1 - \varepsilon), \quad k \rightarrow \infty,$$

then the statement (26) is true, which verifies the asymptotic relation (24) of the Theorem 2. \square

Remark. It is worth mentioning that the asymptotic relation $P(B_k) = P(A_k)(1 + o(1))$ holds if k tends to infinity, which is a direct consequence of the connections (21) and (24).

Remark. In special cases the formula (23) of the Theorem 2. may be computationally applicable for the numerical investigation of the probabilities $P(B_k)$ depending on k , when the generating function of the sequence μ_k has known form.

For instance, if the sequence μ_k equals the Fisher’s logarithmic series (Fisher et al., 1943), which is given by $\mu_k = (\alpha/k)\rho^k$, where μ_k is the expected number of species with k individuals, ρ is a positive number less than 1, and Fisher’s α is a positive constant and it is often used as a measure of biodiversity. In this case we have

$$G(z) = \sum_{i=1}^{\infty} (\alpha/k)\rho^k z^{1/k} = -\alpha \log(1 - \rho z^{1/k}), \quad \mu = \sum_{i=1}^{\infty} (\alpha/k)\rho^k = -\alpha \log(1 - \rho)$$

and consequently

$$\begin{aligned} P(B_k) &= (\alpha/k)\rho^k \int_0^1 \exp \left\{ -\alpha(\log(1 - \rho) + \log(1 - \rho x^{1/k})) \right\} dx = \\ &= (\alpha/k)\rho^k \int_0^1 \exp \left\{ \alpha \log \frac{1 - \rho x^{1/k}}{1 - \rho} \right\} dx = (\alpha/k)\rho^k \int_0^1 \left(\frac{1 - \rho x^{1/k}}{1 - \rho} \right)^\alpha dx. \end{aligned}$$

Another example is the case when the members of the sequence μ_k in the Theorem 2. are proportional to that of a Yule distribution with parameter $\beta > 0$, instead of logarithmic distribution. Let $\mu_k = \alpha\beta \frac{\Gamma(\beta)\Gamma(k)}{\Gamma(k+\beta+1)}$, $k = 1, 2, \dots$ for some $\alpha > 0$. Applying the formula of generating functions of the Yule distributions (see p. 287, Johnson, 2005), then the sequence of probabilities $P(B_k)$ can be formulated as follows

$$P(B_k) = \mu_k \int_0^1 \exp \left\{ -\alpha + \frac{\alpha\beta}{\beta + 1} {}_2F_1[1, 1; \beta + 2; z^{1/k}]z^{1/k} \right\} dx,$$

where ${}_2F_1$ denotes the generalized hypergeometric function.

Conclusions

We have dealt with the model in which a Kendall process describes the evolution of the species after entering the system. The birth and death intensities are assumed to be identical. We have considered inhomogeneous speciation process, for which the speciation rate is bounded, locally integrable and has an asymptotically power type behaviour. This model led (exactly or asymptotically) to Yule abundance distributions instead of a logarithmic one, arising in the homogeneous cases. More precisely, in the inhomogeneous cases the parameters of the models, depending on the speciation rate, are proportional (exactly or asymptotically) to the members of the Yule distribution. This means an enlargement of the class of the possible limit distributions, which can arise for the discrete population dynamical models.

In connection with the sample statistics some results are derived in general and also in special cases (for the logarithmic and Yule distribution), which are related to the random choice of a species or an individual from the whole population of models considered above.

Acknowledgment

The authors are indebted to reviewers for their valuable comments and suggested corrections.

References

- [1] Bowler, M.G. and Kelly, C.K. (2012) On the statistical mechanics of species abundance distributions. *Theoretical Population Biology* 82: 85-91.
- [2] Branson, D. (1991) Inhomogeneous birth-death and birth-death-immigration processes and the logarithmic series distribution. *Stochastic Processes and their Applications* 39, 131-137.
- [3] Branson, D. (2000) Inhomogeneous birth-death and birth-death-immigration processes and the logarithmic series distribution, Part 2. *Stochastic Processes and their Applications*, 86, 183-191.
- [4] Davis, Ph.J. (1972) Gamma Functions and Related Functions. In: *Handbook of Mathematical Functions*, Abramovitz, M. and Stegun, I.A. (Eds.), National Bureau of Standards, USA, Tenth Printing, 253-266.
- [5] Engen, S. and Lande, R. (1996a) Population dynamic models generating the lognormal species abundance distribution. *Mathematical Biosciences*, 132, 169-183.
- [6] Engen, S. and Lande, R. (1996b) Population dynamic models generating species abundance distributions of the gamma type. *Journal of Theoretical Biology*, 178, 325-331.
- [7] Fisher, R. A., Corbet, A. S., Williams, C. B., (1943) The relation between the number of species and the number of individuals in a random sample of an animal population. *J. Animal Ecology*, 12, 42-58.
- [8] Johnson, N.L. Kemp, A.W. Kotz, S. (2005) *Univariate Discrete Distribution*. Third Edition, Wiley & Sohns, New Jersey.
- [9] Karlin, S. and McGregor, J. (1967) The number of mutant forms maintained in a population. *Proc. Fifth Berkeley Symp. Math. Stat. Probab.*, 4, 415-438.
- [10] Kendall, D.G. (1948a) On some models of population growth leading to R. A. Fisher's logarithmic series distribution. *Biometrika*, 35, 6-15.
- [11] Kendall, D.G. (1948b) On the generalized "birth-and-death" process. *Ann Math Stat* 19, 6-15.
- [12] Kingman, J.F.C. (1993) *Poisson Processes*, Clarendon Press, Oxford University.
- [13] Lakatos, L., Szeidl, L. and Telek, M. (2013) *Introduction to Queueing Systems with Telecommunication Applications*, Springer, New York, Heidelberg, Dordrecht, London.

- [14] Lange, K. (2010) *Applied Probability (Second Edition)*, Springer, New York, Dordrecht, Heidelberg, London.
- [15] Newman, M. E. J. Power laws, Pareto distributions and Zipf's law. arXiv:cond-mat/0412004v3, 29 May, 2006, 1-28.
- [16] Prudnikov, A.P., Brychkov, Y.A. and Marichev, O.I. (1998) *Integrals and Series, Vol. 1., Elementary Functions*. Gordon and Breach Sciences Publishers, New York.
- [17] Rolland, J., Condamine, F.L., Jiguet, F., Morlon, H. (2014) Faster speciation and reduced extinction in the tropics contribute to the mammalian latitudinal diversity gradient. *PLOS Biology*, 12, 1, 1-11.
- [18] Simon, H. A. (1955) On a class of skew distribution functions. *Biometrika* 42, 425-440 .
- [19] Watterson, G.A. (1974) Models for the logarithmic species abundance distributions. *Theoret. Population Biol.* 6, 217-250.