Stochastic Analysis of the Extra Clustering Model for Animal Grouping

Michael DRMOTA
Institute for Discrete Mathematics and Geometry
Technical University of Vienna
1040 Vienna
Austria

Michael FUCHS and Yi-Wen LEE
Department of Applied Mathematics
National Chiao Tung University
Hsinchu, 300
Taiwan

November 20, 2014

Abstract

We consider the extra clustering model which was introduced by Durand, Blum and François (2007) in order to describe the grouping of social animals and to test whether genetic relatedness is the main driving force behind the group formation process. Durand and François (2010) provided a first stochastic analysis of this model by deriving (amongst other things) asymptotic expansions for the mean value of the number of groups. In this paper, we will give a much finer analysis of the number of groups. More precisely, we will derive asymptotic expansions for all higher moments and give a complete characterization of the possible limit laws. In the most interesting case (neutral model), we will show a central limit theorem with a surprising normalization. In the remaining cases, the limit law will be either a mixture of a discrete and continuous law or a discrete law. Apart from weak convergence to the limit law, we will also show that all moments converge except in two cases.

1 Introduction and Results

A basic and important problem in biology is to gain an understanding of the dynamics of the group formation process of social animals, which are animals who live in groups, for instance, wolves, gazelles, elephants, lions, etc. In order to solve this problem, biologists have proposed many models for animal grouping, e.g., fusion/fission models, kinship models and models based on game theory; see the introduction of Durand et al. [7] for a detailed discussion.

Key words: Social animals, number of groups, moments, limit laws, singularity perturbation analysis.
2010 Mathematics Subject Classification: 05A16, 60F05, 92B05.

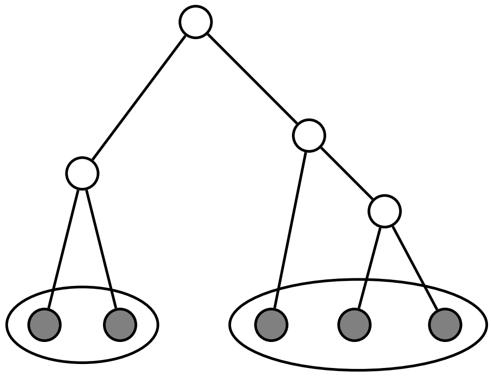

Figure 1: The tree arising from the coalescent process applied to five animals (gray nodes). The number of groups in this example is two.

Some of these models use genetic relatedness as one of the driving forces behind the group formation process. Moreover, in many real-world studies, it has been observed that animals within a group are indeed often genetically related. Thus, in [7], the authors proposed a simplification of previous models where only genetic relatedness is used to decide which animals belong to the same group.

The advantage of such a simplified model is that one can use the coalescent process in order to define group patterns. Moreover, the model is simple enough to devise statistical tests with which one can test whether genetic relatedness really is the major driving force behind the group formation process. For this, the authors in [7] defined the *extra clustering model* which depends on a parameter $0 \leq p \leq 1$. The parameter gives the probability of additional group formation which does not correspond to genetic relatedness. Hence, for $p = 0$, no other factors than genetic relatedness are present and in this case, the authors of [7] called their model the *neutral model*. From a statistical point of view, one is now interested in testing the hypothesis $p = 0$ against $p > 0$. For this purpose, the authors in [7] proposed some statistical tests and applied them to real-world data. The outcome was a good fit of the neutral model for many classes of social animals except classes which have many predators, likely, because in this case, security is another important reason why animals huddle together.

A first probabilistic analysis of the extra clustering model was carried out by Durand and Françis in [8], where they derived asymptotic expansions for the mean number of groups. The main goal of this paper is to provide a much finer stochastic analysis of the extra clustering model. More precisely, we will derive asymptotic expansions of all moments and prove limit distribution results for all values of p.

Before stating our results, we will give a precise definition of the extra clustering model. We start with the case $p = 0$ (neutral model). Here, the model is defined via the Kingman coalescent [19]: start with n animals which are considered to be singleton groups; at every time point pick uniformly at random two groups and let them coalesce; continue this until only one group is left. This random process can be depicted by a rooted, binary tree, where the animals are the leaves and every coalescent event corresponds to the creation of an internal node. If the leaves are drawn at the bottom and the root at the top, then the Kingman coalescent corresponds to a random process building the tree bottom-up; see Figure 1. Alternatively, one can built a random tree top-down as follows: start with the root and two leaves; choose a leave uniformly at random and replace it by an internal node with two leaves; do this until n leaves are created. It is well-known that these two random processes yield the same random model on the set of all rooted, binary trees; see Blum et al. [4]. Moreover, this random model is also equivalent to the Yule-Harding model on phylogenetic trees; see Fuchs and Chang [5] for details.

We recall some properties of the above random tree. First, if the two subtrees of the root have size

j and $n-j$, respectively, then given the size, the two subtrees are again random trees generated by the same model. Moreover, the (random) size of the subtrees is j and $n-j$ with $1 \le j \le n-1$ with equal probabilities, i.e., probability $1/(n-1)$; for these properties see, e.g., [5].

The above random tree was used in [7] to define the random number of groups. More precisely, consider n animals and construct the above random tree. This random tree describes genetic relatedness of the animals. In particular, for a given leaf of the tree, all the animals belonging to the subtree rooted at the father are genetically closely related to the leaf and this set of animals is called a *clade*; see Blum and François [3] and [5]. The number of groups of the n animals is now given by the number of *maximal clades*; see Figure 1. In the sequel, we will denote this number by X_n. From the top-down construction of the random tree and the above stochastic properties, we immediately see that X_n satisfies the following distributional recurrence

$$X_n \stackrel{d}{=} \begin{cases} 1, & \text{if } I_n \in \{1, n-1\}; \\ X_{I_n} + X^*_{n-I_n}, & \text{otherwise,} \end{cases} \qquad (n \ge 3), \tag{1}$$

where $X_2 = 1$, I_n has a uniform distribution on $\{1, \dots, n-1\}$, and X^*_n denotes an independent copy of X_n. This recurrence is explained as follows: the number of groups is computed as the sum of the number of groups of the subtrees of the root unless there is only one maximal clade which is the case if and only if one of the subtrees has size one.

Recurrences of the above type have been extensively studied in the last decades because they also arise in the analysis of certain algorithms and data structures from computer science. In particular, in Hwang and Neininger [16], the authors proposed a very general framework to limit laws of sequences of random variable satisfying distributional recurrence similar to (1). Our above recurrence, although closely related, however, does not fall into the framework of [16]. In particular, some new phenomena not observed before for these recurrences will appear and this makes a detailed analysis of (1) highly interesting.

We next explain the extra clustering model from [7]. As mentioned before, this model depends on a probability p which describes the probability of extra clustering in the group formation process. More precisely, the recurrence (1) for the number of groups is replaced by the following distributional recurrence

$$X_n \stackrel{d}{=} \begin{cases} 1, & \text{with probability } p; \\ 1, & \text{with probability } 1-p \text{ and } I_n \in \{1, n-1\}; \\ X_{I_n} + X^*_{n-I_n}, & \text{with probability } 1-p \text{ and } I_n \notin \{1, n-1\}, \end{cases} \qquad (n \ge 3), \tag{2}$$

where $X_2 = 1$ and notation is as above. Note that $p = 0$ corresponds to the neutral model. For this model, the authors in [8] computed the following asymptotic expansion of the mean

$$\mathbb{E}(X_n) \sim \begin{cases} \dfrac{c(p)}{\Gamma(2(1-p))} n^{1-2p}, & \text{if } 0 \le p < 1/2; \\ \dfrac{\log n}{2}, & \text{if } p = 1/2; \\ \dfrac{p}{2p-1}, & \text{if } 1/2 \le p \le 1, \end{cases}$$

where

$$c(p) = \frac{1}{e^{2(1-p)}} \int_0^1 (1-t)^{-2p} e^{2(1-p)t} \left(1 - (1-p)t^2\right) \mathrm{d}t.$$

The main aim of this paper is to refine this result by proving results for all higher moments and to investigate the limit distribution of X_n for all p.

We start with the neutral model. Here, the result of [8] can be generalized as follows.

Proposition 1 ($p = 0$). *We have,*

$$\mathrm{Var}(X_n) \sim \frac{(1-e^{-2})^2}{4} n \log n$$

and for all $k \geq 3$,

$$\mathbb{E}\left(X_n - \mathbb{E}(X_n)\right)^k \sim (-1)^k \frac{2k}{k-2} \left(\frac{1-e^{-2}}{4}\right)^k n^{k-1}.$$

Remark 1. Note that from this result, it follows that the limit law of X_n cannot be found from the method of moments; for the latter method see Section 30 in Billingsley [2]. This is in sharp contrast to [16], where the method of moments was applied to many examples of X_n satisfying a recurrence similar to (1).

In order to find the limit law, we will use an approach based on singularity perturbation analysis; for details see Section 3. Applying this approach gives the following surprising central limit theorem.

Theorem 1 ($p = 0$). *We have,*

$$\frac{X_n - \mathbb{E}(X_n)}{\sqrt{\mathrm{Var}(X_n)/2}} \xrightarrow{d} N(0,1),$$

where $N(0,1)$ *denotes the standard normal distribution.*

The surprising fact about this result is the curious normalization by half of the variance. Note that a similar (but seemingly unrelated) phenomenon was also observed by Janson and Kersting [18] in their analysis of the total external path length of the Kingman coalescent.

Remark 2. A probabilistic proof of the above central limit theorem shedding further light on the curious normalization was given recently by Janson; see [17].

We next consider the case $0 < p < 1/2$. Here, the result for the moments reads as follows.

Proposition 2 ($0 < p < 1/2$). *For all* $k \geq 1$,

$$\mathbb{E}X_n^k \sim \frac{d_k}{\Gamma(k(1-2p)+1)} n^{k(1-2p)},$$

where d_k *is recursively given by* $d_1 = c(p)$ *and for* $k \geq 2$

$$d_k = \frac{1-p}{(k-1)(1-2p)} \sum_{j=1}^{k-1} \binom{k}{j} d_j d_{k-j}.$$

In contrast to the case $p = 0$, now the method of moments can be applied and yields the limiting distribution of X_n. Moreover, this limit result can also be proved with our proof method of Theorem 1 based on singularity perturbation analysis. This second approach in addition gives information about the distribution function of the limit law.

Theorem 2 ($0 < p < 1/2$). *We have,*

$$\frac{X_n}{n^{1-2p}} \xrightarrow{d} X$$

with convergence of all moments. Here, X *is a random variable whose moment-generating function is given by*

$$\mathbb{E}\left(e^{yX}\right) = \frac{1}{2\pi i} \int_{\mathcal{H}} \Phi(y,v) e^{-v} \mathrm{d}v,$$

where $\mathcal{H}$ is the Hankel contour starting in the upper half plane at $+\infty$ and winding around 0 counter-clockwise before tending to $+\infty$ in the lower half plane and

$$\Phi(y,v)=\frac{4(1-2p)^2-ypm(p)4^p(1-p)^{2p-1}v^{2p-1}}{4(1-2p)^2v-ym(p)4^p(1-p)^{2p}v^{2p}}$$

with determination of the powers in t chosen such that the branch cut is at $[0,\infty)$ and

$$m(p)=\frac{M_{p,(1-2p)/2}(-2(1-p))}{W_{p,(1-2p)/2}(-2(1-p))},$$

where $M_{\kappa,\mu}(z)$ and $W_{\kappa,\mu}(z)$ are the Whittaker M and Whittaker W functions (see Beals and Wong [1]).

From the above explicit form of the moment-generating function of the limit law, we obtain the following consequence.

Corollary 1 ($0<p<1/2$)**.** *The distribution function of X is the sum of of a discrete distribution of measure $p/(1-p)$ that is concentrated at 0 and a continuous distribution on $[0,\infty)$ with density*

$$f(x)=-\delta(p)\frac{1-2p}{1-p}\sum_{k\geq 0}\frac{\delta(p)^k}{k!\Gamma(2(k+1)p-k)}x^k, \tag{3}$$

where

$$\delta(p)=\frac{(1-2p)^2}{e^{2\pi ip}4^{p-1}(1-p)^{2p}m(p)}.$$

Remark 3. The density from the above corollary can be computed numerically for each $p<1/2$. Furthermore, for $p=1/4$, one obtains

$$f(x)=0.3780064347\cdots e^{-0.2525054668\cdots x^2}.$$

For other values of p the resulting expressions are usually less explicit; see Figure 2 for a plot of the density functions for several values of p.

The next case we consider is $p=1/2$. Here, we again have a result for all higher moments.

Proposition 3 ($p=1/2$)**.** *We have,*

$$\mathbb{E}(X_n^k)\sim\frac{k!J_{2k-1}}{(2k-1)!2^{2k-1}}\log^{2k-1}n,$$

where J_{2k-1} are the Euler numbers of odd degree (see, e.g., page 144 in Flajolet and Sedgewick [15]).

Thus, similar to $p=0$, the method of moments cannot be applied in this situation. However, singular perturbation analysis works and yields the following limit distribution result.

Theorem 3 ($p=1/2$)**.** *We have,*

$$X_n\xrightarrow{d}X,$$

where X has a discrete law given by $P(X=k)=2^{-k}/(2k-1)\binom{2k}{k}$, $k\geq 1$, which has the following explicit characteristic function

$$\mathbb{E}\left(e^{itX}\right)=1-\sqrt{1-e^{it}}.$$

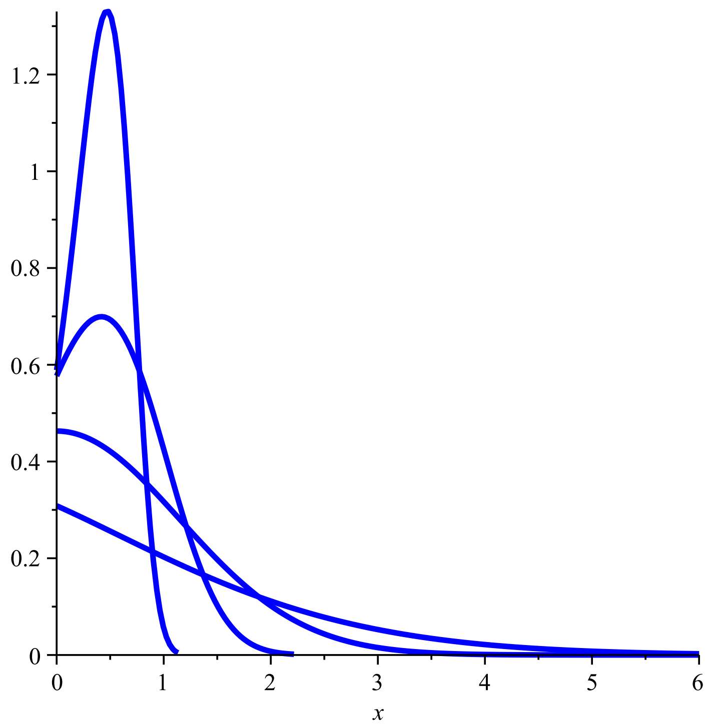

Figure 2: The density $f(x)$ of the continuous part of the distribution of X for $p = 1/8, 3/16, 1/4, 5/16$ (top to bottom).

Finally, we turn to the case $1/2 < p \leq 1$. Here, the result for the moments is as follows.

Proposition 4 $(1/2 < p \leq 1)$**.** *For all* $k \geq 1$*,*

$$\mathbb{E}(X_n^k) \sim e_k,$$

where e_k *is recursively given by* $e_1 = p/(2p-1)$ *and for* $k \geq 2$

$$e_k = \frac{1-p}{2p-1} \sum_{j=1}^{k-1} \binom{k}{j} e_j e_{k-j} + \frac{p}{2p-1}.$$

Thus, again the method of moments can be used to derive the limiting distribution (alternatively, the result can also be proved by singularity perturbation analysis).

Theorem 4 $(1/2 < p < 1)$**.** *We have,*

$$X_n \xrightarrow{d} X$$

with convergence of all moment. Here, X *has a discrete law given by* $P(X = k) = p^k(1-p)^{k-1}/(2(2k-1))\binom{2k}{k}$*,* $k \geq 1$*, which has the following explicit moment-generating function*

$$\mathbb{E}\left(e^{yX}\right) = \frac{1 - \sqrt{1 - 4p(1-p)e^y}}{2(1-p)}.$$

Remark 4. For $p = 1$, the above result is also valid if one takes the limit of the above expression for the moment-generating function which gives the (trivial) limit law $X = 1$ a.s..

Remark 5. Note that Theorem 3 and Theorem 4 can be merged. However, note that there is an important difference between these results: in Theorem 3 we only have convergence in distribution, whereas in Theorem 4 also all moments do converge.

Overall, our above results combined give a full picture of the limiting behavior of the number of groups under the extra clustering model. In particular, we see that the limit law is continuous for $p = 0$, is a mixture of a discrete and continuous distribution for $0 < p < 1/2$, and finally becomes discrete as $1/2 \leq p \leq 1$ (and degenerates at $p = 1$). Moreover, from a mathematical point of view, an interesting aspect is that the limit law can be obtained via the method of moments if and only if $0 < p < 1/2$ and $1/2 < p \leq 1$, but not in the two cases $p = 0$ and $p = 1/2$.

We conclude the introduction with a short sketch of the paper. In the next section, we will consider moments and prove Propositions 1-4. As a method of proof, we will use singularity analysis which is a standard tool in analytic combinatorics (see the next section for details). In Section 3, we will introduce the tools needed for the proofs of Theorem 1-4. More precisely, this section will contain a short discussion of Whittaker functions and some of their properties which are needed in the proofs. Moreover, we will explain our approach to limit laws via singularity perturbation analysis. The proofs of the limit distribution results will then be contained in Section 4. We will end the paper with a conclusion.

2 Moments

In this section, we will investigate the moments of X_n satisfying (2) for all values of p. The method we use for this has already been used in many other studies and was nicknamed "moment pumping"; see, e.g., Chern et al. [6] or Fill and Kapur [12] and references therein. It is based on induction and singularity analysis. The latter is a standard tool of analytic combinatorics, see Chapter VI of [15], and says – in a nutshell – that the leading asymptotic behavior of the coefficients a_n of a power series $f(z) = \sum_{n\geq 0} a_n z^n$ is mainly governed by the kind of the dominating singularity z_0 of $f(z)$ on the radius of convergence $|z| = |z_0| = R$. For example, if $z_0 = 1$ and we have

$$f(z) = A(1-z)^{\alpha} + \mathcal{O}((1-z)^{\beta})$$

for $z \to 1$, $z \in \Delta$, where α, β are real numbers and Δ is a so-called Δ-domain of the form

$$\Delta = \{z \in \mathbb{C} : |z| < 1 + \delta, \, |\arg(z-1)| > \varphi\}, \qquad (\delta > 0, 0 < \varphi < \pi/2),$$

then we have

$$a_n = A\frac{n^{-\alpha-1}}{\Gamma(-\alpha)} + \mathcal{O}\left(n^{\max\{-\beta-1,-\alpha-2\}}\right).$$

Actually, we can also work without an error term. For example, $f(z) \sim A(1-z)^{\alpha}$ as $z \to 1$, $z \in \Delta$, implies $a_n \sim An^{-\alpha-1}/\Gamma(-\alpha)$. This tool will be extensively used below.

We next explain in more details the above mentioned method of moment-pumping. Therefore, we consider the moment-generating function of X_n which by (2) satisfies the recurrence

$$\mathbb{E}\left(e^{yX_n}\right) = pe^y + (1-p)\frac{2}{n-1}e^y + \frac{1-p}{n-1}\sum_{j=2}^{n-2}\mathbb{E}\left(e^{yX_j}\right)\mathbb{E}\left(e^{yX_{n-j}}\right), \qquad (n \geq 3)$$

with initial condition $\mathbb{E}\left(e^{yX_2}\right) = e^y$. Next, set

$$X(y,z) = \sum_{n\geq 2}\mathbb{E}\left(e^{yX_n}\right)z^n.$$

Then, by a straightforward computation

$$z\frac{\partial}{\partial z}X(y,z) = X(y,z) + (1-p)X(y,z)^2 + e^y\frac{z^2(1-(1-p)z^2)}{(1-z)^2} \tag{4}$$

with $X(y,0) = 0$.

From this, differential equations for generating functions of moments of X_n can be obtained by differentiating with respect to y and setting $y = 0$. In particular, the arising differential equations are all of the following general form

$$\tilde{f}'(z) = \left(\frac{1}{z} + \frac{2(1-p)z}{1-z}\right) f(z) + g(z), \tag{5}$$

where $g(z)$ is a function of generating functions of moments of smaller order. Thus, we have a recursive scheme with which generating functions of moments can be computed inductively once a general solution of the above differential equation is known.

Lemma 1. *Let $f(z)$ and $g(z)$ be functions which are analytic at zero and satisfy*

$$f'(z) = \left(\frac{1}{z} + \frac{2(1-p)z}{1-z}\right) f(z) + g(z),$$

where $f(0) = 0$. Then,

$$f(z) = \frac{z}{(1-z)^{2(1-p)}e^{2(1-p)z}}\int_0^z \frac{(1-t)^{2(1-p)}e^{2(1-p)t}}{t} g(t)\mathrm{d}t.$$

Proof. This is proved by applying the standard approach for solving first-order differential equations. ∎

We will use this general solution and induction to obtain the singularity expansion (in a Δ-domain) of generating functions of moments of all order. Moreover, in the same way, generating functions of moments are also proved to be analytic in a suitable domain. Both these properties will follow from closure properties of singularity analysis; see Fill et al. [11] or Section VI.10 in [15].

We will first demonstrate how this works for the neutral model.

Proof of Proposition 1: $p = 0$. We start with mean and variance. Differentiating (4) with $p = 0$ once and twice and setting $y = 0$ gives

$$M'(z) = \left(\frac{1}{z} + \frac{2z}{1-z}\right) M(z) + \frac{z(1+z)}{1-z}$$

and

$$S'(z) = \left(\frac{1}{z} + \frac{2z}{1-z}\right) S(z) + \frac{2}{z}M(z)^2 + \frac{z(1+z)}{1-z},$$

with the notation

$$M(z) = \sum_{n\geq 2}\mathbb{E}(X_n)z^n, \qquad S(z) = \sum_{n\geq 2}\mathbb{E}(X_n^2)z^n.$$

Now, for the mean, an application of Lemma 1 gives

$$M(z) = \frac{(-1+e^{2z}+2ze^{2z}-2z^2e^{2z})z}{(1-z)^2 4e^{2z}}. \tag{6}$$

Thus, as $z \to 1$, $z \in \Delta$,

$$M(z) = \frac{1-e^{-2}}{4} \cdot \frac{1}{(1-z)^2} + \mathcal{O}\left(\frac{1}{1-z}\right).$$

Consequently, by applying singularity analysis,

$$\mathbb{E}(X_n) = \frac{1-e^{-2}}{4} n + \mathcal{O}(1).$$

Next, for the second moment, again by Lemma 1

$$S(z) = \frac{z}{(1-z)^2 e^{2z}} \int_0^z \left(\frac{2(1-t)^2 e^{2t}}{t^2} M(t)^2 + (1-t)^2 e^{2t} \right) \mathrm{d}t.$$

Using (6) together with a proper use of a computer algebra system, we obtain that for the integrand, as $t \to 1$, $t \in \Delta$,

$$\frac{2(1-t)^2 e^{2t}}{t^2} M(t)^2 + (1-t)^2 e^{2t} \sim \frac{(e^2-1)^2}{8e^2} \cdot \frac{1}{(1-t)^2} + \frac{(e^2-1)^2}{4e^2} \cdot \frac{1}{1-t}.$$

This leads to,

$$S(z) = \frac{(1-e^{-2})^2}{8} \cdot \frac{1}{(1-z)^3} + \frac{(1-e^{-2})^2}{4} \cdot \frac{1}{(1-z)^2} \log\left(\frac{1}{1-z}\right) + o\left(\frac{1}{(1-z)^2} \log\left(\frac{1}{1-z}\right)\right)$$

as $z \to 1$, $z \in \Delta$. Hence, again by singularity analysis,

$$\mathbb{E}(X_n^2) = \frac{(1-e^{-2})^2}{16} n^2 + \frac{(1-e^{-2})^2}{4} n \log n + o(n \log n).$$

From this and the above expansion of the mean, we obtain that

$$\mathrm{Var}(X_n) \sim \frac{(1-e^{-2})^2}{4} n \log n.$$

From the last two results, we also obtain a strong law of large numbers for X_n (with the coupling arising from the top-down construction of the random tree from the previous section).

Theorem 5 ($p = 0$)**.** *We have,*

$$P\left(\lim_{n\to\infty} \left|\frac{X_n}{\mathbb{E}(X_n)} - 1\right|\right) = 1.$$

In other words,

$$X_n \sim \mathbb{E}(X_n) \qquad \textit{a.s.}.$$

Proof. First, consider $n = k^2$. Then, by Chebyshev's inequality,

$$P\left(\left|\frac{X_{k^2}}{\mathbb{E}(X_{k^2})} - 1\right| \geq \epsilon\right) = P(|X_{k^2} - \mathbb{E}(X_{k^2})| \geq \epsilon\mathbb{E}(X_{k^2})) = \mathcal{O}\left(\frac{\log k}{k^2}\right)$$

for all $\epsilon > 0$, where in the last step, we used the above results for the mean and variance of X_n. A standard application of the lemma of Borel-Cantelli now gives

$$\frac{X_{k^2}}{\mathbb{E}(X_{k^2})} \xrightarrow{\text{a.s.}} 1 \qquad (k \to \infty). \tag{7}$$

Next, for general n, find k such that

$$k^2 \le n < (k+1)^2.$$

Note that by the above asymptotics for the mean, we have that

$$\mathbb{E}(X_{(k+1)^2}) \sim \mathbb{E}(X_{k^2}) \qquad (k \to \infty). \tag{8}$$

Moreover, the fact that X_n is non-decreasing (from the coupling) gives

$$\frac{X_{k^2}}{\mathbb{E}(X_{(k+1)^2})} \le \frac{X_n}{\mathbb{E}(X_n)} \le \frac{X_{(k+1)^2}}{\mathbb{E}(X_{k^2})}.$$

From this, the claimed results follows by using (7) and (8). ∎

This result suggests to look at central moments. Hence, we set

$$\bar{X}(y,z) := X(y, ze^{-ya}) = \sum_{n\ge 2} \mathbb{E}\left(e^{y(X_n - an)}\right) z^n$$

with $a := (1-e^{-2})/4$. Then, (4) becomes (recall that $p = 0$)

$$z\frac{\partial}{\partial z}\bar{X}(y,z) = \bar{X}(y,z) + \bar{X}^2(y,z) + e^{y(1-2a)}z^2 + \frac{2e^{y(1-3a)}z^3}{1-ze^{-ya}}.$$

Now, taking the k-th derivative with respect to y and setting $y = 0$ yields for

$$\bar{M}^{[k]}(z) := \frac{\partial^k}{\partial y^k}\bar{X}(y,z)\Big|_{y=0}$$

the differential equation

$$\bar{M}^{[k]\prime}(z) = \left(\frac{1}{z} + \frac{2z}{1-z}\right)\bar{M}^{[k]}(z) + \frac{1}{z}\sum_{j=1}^{k-1}\binom{k}{j}\bar{M}^{[j]}(z)\bar{M}^{[k-j]}(z) + \bar{h}^{[k]}(z),$$

where $\bar{M}^{[k]}(0) = 0$ and

$$\bar{h}^{[k]}(z) = (1-2a)^k z + \frac{\mathrm{d}^k}{\mathrm{d}y^k}\frac{2e^{y(1-3a)}z^2}{1-ze^{-ya}}\Big|_{y=0}.$$

This differential equation is of the type (5). Thus, we can apply Lemma 1 and induction to obtain the following lemma.

Lemma 2. *For $k \ge 3$, as $z \to 1$, $z \in \Delta$,*

$$\bar{M}^{[k]}(z) \sim \frac{2(-1)^k k! a^k}{(k-2)(1-z)^k}.$$

Proof. First note that from the computations above for the mean and the variance, we have the following bounds, as $z \to 1$, $z \in \Delta$,

$$\bar{M}^{[1]}(z) = \mathcal{O}\left(\frac{1}{1-z}\right) \qquad \text{and} \qquad \bar{M}^{[2]}(z) = \mathcal{O}\left(\frac{1}{(1-z)^2}\log\frac{1}{1-z}\right). \tag{9}$$

We prove our claim by induction. Since the proofs for the base step and the induction step are the same, we merge them into one. So, assume that the claim holds for all $k' < k$. In order to show that it holds for k, we use Lemma 1 which yields

$$\bar{M}^{[k]}(z) = \frac{z}{(1-z)^2 e^{2z}} \int_0^z \frac{(1-t)^2 e^{2t}}{t} \left(\frac{1}{t} \sum_{j=1}^{k-1} \binom{k}{j} \bar{M}^{[j]}(t) \bar{M}^{[k-j]}(t) + \bar{h}^{[k]}(t) \right) \mathrm{d}t. \tag{10}$$

We first consider the two terms inside the bracket. For the first, by (9) and induction hypothesis, we obtain that, as $t \to 1$, $t \in \Delta$,

$$\frac{1}{t} \sum_{j=1}^{k-1} \binom{k}{j} \bar{M}^{[j]}(t) \bar{M}^{[k-j]}(t) = \mathcal{O}\left(\frac{1}{(1-t)^{k+\epsilon}} \right),$$

where $\epsilon > 0$ is an arbitrarily small constant (this constant comes from the additional log term of $\bar{M}^{[2]}(z)$). For the second term, it is not complicated to see that, as $t \to 1$, $t \in \Delta$,

$$\bar{h}^{[k]}(t) \sim \frac{2(-1)^k a^k}{(1-t)^{k+1}}.$$

Thus, for the integrand of (10), as $t \to 1$, $t \in \Delta$,

$$\frac{(1-t)^2 e^{2t}}{t} \left(\frac{1}{t} \sum_{j=1}^{k-1} \binom{k}{j} \bar{M}^{[j]}(t) \bar{M}^{[k-j]}(t) + \bar{h}^{[k]}(t) \right) \sim \frac{2e^2 (-1)^k a^k}{(1-t)^{k-1}}.$$

Hence, by the closure properties of singularity analysis, as $z \to 1$, $z \in \Delta$,

$$\int_0^z \frac{(1-t)^2 e^{2t}}{t} \left(\frac{1}{t} \sum_{j=1}^{k-1} \binom{k}{j} \bar{M}^{[j]}(t) \bar{M}^{[k-j]}(t) + \bar{h}^{[k]}(t) \right) \mathrm{d}t \sim \frac{2e^2 (-1)^k a^k}{(k-2)(1-z)^{k-2}}.$$

Inserting this into (10) gives the claimed result. ∎

The proposed expansion for all central moments of order higher than two in Proposition 1 now follows from Lemma 2 and singularity analysis. In particular, note that by

$$\mathbb{E}(X_n - \mathbb{E}(X_n))^k = \mathbb{E}(X_n - an)^k + \mathcal{O}\left(|\mathbb{E}(X_n - an)^{k-1}| \right),$$

we can easily transfer asymptotic results for $\mathbb{E}(X_n - an)^k$ to corresponding ones for $\mathbb{E}(X_n - \mathbb{E}(X_n))^k$.

Proof of Proposition 2: $0 < p < 1/2$. Here, in contrast to before, we directly consider moments instead of central moments. Therefore, we set

$$M^{[k]}(z) := \frac{\partial^k}{\partial y^k} X(y, z) \Big|_{y=0} = \sum_{n \geq 2} \mathbb{E} X_n^k z^n.$$

Then, (4) implies that

$$M^{[k]\prime}(z) = \left(\frac{1}{z} + \frac{2(1-p)z}{1-z} \right) M^{[k]}(z) + \frac{1-p}{z} \sum_{j=1}^{k-1} \binom{k}{j} M^{[j]}(z) M^{[k-j]}(z) + h(z).$$

where $M^{[k]}(0) = 0$ and

$$h(z) = \frac{z(1-(1-p)z^2)}{(1-z)^2}.$$

By Lemma 1, the solution of this differential equation is given by

$$M^{[k]}(z) = \frac{z}{(1-z)^{2(1-p)}e^{2(1-p)z}} \int_0^z \frac{(1-t)^{2(1-p)}e^{2(1-p)t}}{t} \left(\frac{1-p}{t} \sum_{j=1}^{k-1} \binom{k}{j} M^{[j]}(t) M^{[k-j]}(t) + h(t) \right) \mathrm{d}t. \tag{11}$$

Now, the proof of Proposition 2 follows from the next lemma by singularity analysis.

Lemma 3. *For $k \geq 1$, as $z \to 1$, $z \in \Delta$,*

$$M^{[k]}(z) \sim \frac{d_k}{(1-z)^{k(1-2p)+1}}.$$

Proof. We start with $k = 1$. Here, according to (11), we have

$$M^{[1]}(z) = \frac{z}{(1-z)^{2(1-p)}e^{2(1-p)z}} \int_0^z \frac{(1-t)^{2(1-p)}e^{2(1-p)t}}{t} h(t) \mathrm{d}t.$$

Note that the integrand has the singularity expansion, as $t \to 1$, $t \in \Delta$,

$$\frac{(1-t)^{2(1-p)}e^{2(1-p)t}}{t} h(t) \sim \frac{pe^{2(1-p)}}{(1-t)^{2p}}.$$

Since $2p < 1$, applying the closure properties of singularity analysis yields, as $z \to 1$, $z \in \Delta$,

$$\int_0^z \frac{(1-t)^{2(1-p)}e^{2(1-p)t}}{t} h(t) \mathrm{d}t \sim e^{2(1-p)} d_1.$$

Inserting this into the above expression for $M^{[1]}(z)$ gives the claimed asymptotics for $k = 1$.

Now, assume that the claim is true for all $k' < k$. We want to show that it also holds for k. First, observe that by the induction hypothesis, the integrand of (11) satisfies, as $t \to 1$, $t \in \Delta$,

$$\begin{aligned} &\frac{(1-t)^{2(1-p)}e^{2(1-p)t}}{t} \left(\frac{1-p}{t} \sum_{j=1}^{k-1} \binom{k}{j} M^{[j]}(t) M^{[k-j]}(t) + h(t) \right) \\ &\qquad \sim (1-p)e^{2(1-p)} \left(\sum_{j=1}^{k-1} \binom{k}{j} d_j d_{k-j} \right) \cdot \frac{1}{(1-t)^{(k-1)(1-2p)-1}}. \end{aligned}$$

Thus, by the closure properties of singularity analysis, as $z \to 1$, $z \in \Delta$,

$$\int_0^z \frac{(1-t)^{2(1-p)}e^{2(1-p)t}}{t} \left(\frac{1-p}{t} \sum_{j=1}^{k-1} \binom{k}{j} M^{[j]}(t) M^{[k-j]}(t) + h(t) \right) \mathrm{d}t \sim \frac{e^{2(1-p)} d_k}{(1-z)^{(k-1)(1-2p)}}.$$

Inserting this into (11) gives, as $z \to 1$, $z \in \Delta$,

$$M^{[k]}(z) \sim \frac{d_k}{(1-z)^{k(1-2p)+1}}$$

which is the claimed result. ∎

Proof of Proposition 3: $p = 1/2$. Here, we again work with the moments instead of central moments and the proof is similar as in the previous case. More precisely, we show the following lemma.

Lemma 4. *For $k \geq 1$, as $z \to 1$, $z \in \Delta$,*

$$M^{[k]}(z) \sim \frac{b_k}{1-z} \log^{2k-1} \frac{1}{1-z},$$

where b_k is recursively given by $b_1 = 1/2$ and for $k \geq 2$

$$b_k = \frac{1}{2(2k-1)} \sum_{j=1}^{k-1} \binom{k}{j} b_j b_{k-j}. \tag{12}$$

Proof. The proof is again by induction. We start with $k = 1$. In this case, the integrand of (11) satisfies, as $t \to 1$, $t \in \Delta$,

$$\frac{(1-t)e^t}{t} h(t) \sim \frac{e}{2(1-t)}.$$

Thus, as $z \to 1$, $z \in \Delta$,

$$\int_0^z \frac{(1-t)e^t}{t} h(t) \mathrm{d}t \sim \frac{e}{2} \cdot \log \frac{1}{1-z}.$$

Inserting this into (11) gives the claimed result.

For the induction step, assume that the claim holds for all $k' < k$. Then, for the proof that it also holds for k, by the induction hypothesis, the integrand of (11) satisfies, as $t \to 1$, $t \in \Delta$,

$$\frac{(1-t)e^t}{t} \left(\frac{1}{2t} \sum_{j=1}^{k-1} \binom{k}{j} M^{[j]}(t) M^{[k-j]}(t) + h(t) \right) \sim \frac{e}{2} \left(\sum_{j=1}^{k-1} \binom{k}{j} b_j b_{k-j} \right) \cdot \frac{1}{1-t} \log^{2k-2} \frac{1}{1-t}.$$

Hence, by the closure properties of singularity analysis, we obtain that, as $z \to 1$, $z \in \Delta$,

$$\int_0^z \frac{(1-t)e^t}{t} \left(\frac{1}{2t} \sum_{j=1}^{k-1} \binom{k}{j} M^{[j]}(t) M^{[k-j]}(t) + h(t) \right) \mathrm{d}t \sim e b_k \log^{2k-1} \frac{1}{1-z}.$$

Inserting this into (11) concludes the induction step. ∎

The proof of Proposition 3 follows from this by singularity analysis and the following lemma.

Lemma 5. *The solution of the recurrence (12) from Lemma 4 is given by*

$$b_k = \frac{k! J_{2k-1}}{(2k-1)! 2^{2k-1}},$$

where J_{2k-1} are the Euler numbers of odd degree.

Proof. We use generating functions. Set

$$B(z) = \sum_{k \geq 1} b_k \frac{z^k}{k!}.$$

Then, the recurrence (12) becomes

$$B(z)B'(z) = 2zB''(z) + B'(z) - 1/2$$

with $B(0) = 0$. Integrating yields

$$4zB'(z) = B(z)^2 + 2B(z) + z$$

which has the solution

$$B(z) = \sqrt{z} \tan\left(\frac{\sqrt{z}}{2}\right).$$

Expanding provides the claimed result. ∎

Proof of Proposition 4. This final case is again treated similar as the two previous cases. More precisely, the result follows from the following lemma and singularity analysis.

Lemma 6. *For $k \geq 1$, as $z \to 1$, $z \in \Delta$,*

$$M^{[k]}(z) \sim \frac{e_k}{1-z}.$$

Proof. Again, we use induction and (11). First, for $k = 1$, the integrand of (11) satisfies, as $t \to 1$, $t \in \Delta$,

$$\frac{(1-t)^{2(1-p)}e^{2(1-p)t}}{t}h(t) \sim \frac{pe^{2(1-p)}}{(1-t)^{2p}}.$$

Then, from the closure properties of singularity analysis, as $z \to 1$, $z \in \Delta$,

$$\int_0^z \frac{(1-t)^{2(1-p)}e^{2(1-p)t}}{t}h(t)\mathrm{d}t \sim \frac{pe^{2(1-p)}}{(2p-1)(1-z)^{2p-1}}.$$

Inserting this into (11) gives the claimed result.

Next, assume by induction that the claim holds for all $k' < k$. In order to show it for k note that the integrand of (11) satisfies, as $t \to 1$, $t \in \Delta$,

$$\frac{(1-t)^{2(1-p)}e^{2(1-p)t}}{t}\left(\frac{1-p}{t}\sum_{j=1}^{k-1}\binom{k}{j}M^{[j]}(t)M^{[k-j]}(t) + h(t)\right)$$
$$\sim e^{2(1-p)}\left((1-p)\sum_{j=1}^{k-1}\binom{k}{j}e_j e_{k-j} + p\right) \cdot \frac{1}{(1-t)^{2p}}.$$

Thus, by the closure properties of singularity analysis, we obtain that, as $z \to 1$, $z \in \Delta$,

$$\int_0^z \frac{(1-t)^{2(1-p)}e^{2(1-p)t}}{t}\left(\frac{1-p}{t}\sum_{j=1}^{k-1}\binom{k}{j}M^{[j]}(t)M^{[k-j]}(t) + h(t)\right) \sim \frac{e^{2(1-p)}e_k}{(1-z)^{2p-1}}.$$

Inserting this into (11) gives the desired result. ∎

Since we have now considered all cases for p, our analysis of the moments is complete.

3 Whittaker Functions and Singularity Perturbation Analysis

In this section, we will explain our analytic method used for proving Theorem 1-4. The method will rely on the explicit solution of (4) which will involve Whittaker functions. Thus, properties of Whittaker functions will play a crucial role and we will recall them below. The method itself then uses singularity perturbation analysis and will also be explained in details below.

We start by solving (4).

Solution of (4). Note that (4) is a Riccati differential equation for which a standard solution procedure exists. Therefore, set

$$\tilde{X}(y,z) = \frac{X(y,z)}{z}.$$

Then, (4) becomes

$$\frac{\partial}{\partial z}\tilde{X}(y,z) = (1-p)\tilde{X}(y,z)^2 + e^y \frac{1-(1-p)z^2}{(1-z)^2}$$

with $\tilde{X}(y,0) = 0$. Next, set

$$\tilde{X}(y,z) = -\frac{1}{1-p} \cdot \frac{V'(y,z)}{V(y,z)},$$

where $V(y,0) = 1$ and differentiation is with respect z. Then, we obtain the second-order differential equation

$$V''(y,z) + (1-p)e^y \frac{1-(1-p)z^2}{(1-z)^2} V(y,z) = 0$$

with $V(y,0) = 1$ and $V'(y,0) = 0$. This differential equation is a variant of Whittaker's differential equation. Thus, its solution can be expressed in terms of the Whittaker functions as follows

$$\begin{aligned} V(y,z) = {} & M_{-(1-p)e^{y/2}, \sqrt{1-4p(1-p)e^y}/2}\left(2(1-p)e^{y/2}(z-1)\right) \\ & + c(y) W_{-(1-p)e^{y/2}, \sqrt{1-4p(1-p)e^y}/2}\left(2(1-p)e^{y/2}(z-1)\right) \end{aligned}$$

with

$$c(y) = \frac{\left(1 + \sqrt{1-4p(1-p)e^y} - 2(1-p)e^{y/2}\right) M_{-(1-p)e^{y/2}+1, \sqrt{1-4p(1-p)e^y}/2}\left(-2(1-p)e^{y/2}\right)}{2W_{-(1-p)e^{y/2}+1, \sqrt{1-4p(1-p)e^y}/2}\left(-2(1-p)e^{y/2}\right)}.$$

We will work in the next section with this explicit solution. Consequently, we will need some background knowledge on Whittaker functions which we will recall next.

Whittaker Functions. Here, we gather some properties of the Whittaker functions. The exposition will follow [1]. We start with the definition of the Whittaker functions which are independent solutions of Whittaker's differential equation

$$v''(z) + \left(-\frac{1}{4} + \frac{\kappa}{z} + \frac{1-4\mu^2}{4z^2}\right) v(z) = 0.$$

They can be expressed as follows

$$\begin{aligned} M_{\kappa,\mu}(z) &= e^{-z/2} z^{\mu+1/2} M\left(\mu - \kappa + \frac{1}{2}, 1+2\mu, z\right), \\ W_{\kappa,\mu}(z) &= e^{-z/2} z^{\mu+1/2} U\left(\mu - \kappa + \frac{1}{2}, 1+2\mu, z\right). \end{aligned}$$

Here, $M(a,c;z)$ and $U(a,c;z)$ are the Kummer functions. The former is defined for all $a, c, z \in \mathbb{C}$ with $c \neq 0, -1, -2, \ldots$ by the following series

$$M(a,c;z) = \sum_{\ell=0}^{\infty} \frac{(a)_\ell}{(c)_\ell \ell!} z^\ell,$$

where $(a)_\ell$ is the Pochhammer symbol

$$(a)_\ell := a(a+1)\cdots(a+\ell-1).$$

Note that the above expression shows that $M(a,c;z)$ is analytic in all three variables. The definition of the Kummer function of second kind, namely $U(a,c;z)$, is slightly more involved. More precisely, $U(a,c;z)$ is defined as

$$U(a,c;z) = \frac{\Gamma(1-c)}{\Gamma(a+1-c)}M(a,c;z) + \frac{\Gamma(c-1)}{\Gamma(a)}z^{1-c}M(a+1-c,2-c;z)$$

for all a, c, z with $c \notin \mathbb{Z}$. The definition can be extended to $c = m \in \mathbb{N}$ (where the limit exists) as follows

$$\begin{aligned} U(a,m;z) = \frac{(-1)^m}{\Gamma(a+1-m)(m-1)!}\Big(&M(a,m;z)\log z \\ &+ \sum_{\ell=0}^{\infty}\frac{(a)_\ell}{(m)_\ell \ell!}\left(\psi(a+\ell)-\psi(\ell+1)-\psi(m+\ell)\right)z^\ell\Big) \\ &+ \frac{(m-2)!}{\Gamma(a)}z^{1-m}\sum_{\ell=0}^{m-2}\frac{(a+1-m)_\ell}{(2-m)_\ell \ell!}z^\ell \end{aligned}$$

with $\psi(z) = \Gamma'(z)/\Gamma(z)$. We will in the sequel choose the determination of log and powers such that we have a branch cut at $[0,\infty)$. Then, from the above definitions we obtain the following lemma.

Lemma 7. *Assume that $\mu \neq -1/2, -1, -3/2, \cdots$. Then, both of the Whittaker functions are analytic on $\mathbb{C} \setminus [0,\infty)$.*

Finally, note that the above expressions also give singularity expansions as $z \to 0$. For instance, if $\mu \neq -1/2, -1, -3/2, \ldots$, then

$$M_{\kappa,\mu}(z) \sim z^{\mu+1/2}, \qquad (z \to 0)$$

and if in addition $\mu \neq 0, 1/2, 1, \cdots$, then

$$W_{\kappa,\mu}(z) \sim \begin{cases} \dfrac{\Gamma(2\mu)}{\Gamma(\mu-\kappa+1/2)}z^{-\mu+1/2}, & \text{if } \mu > 0; \\ \dfrac{\Gamma(-2\mu)}{\Gamma(-\mu-\kappa+1/2)}z^{\mu+1/2}, & \text{if } \mu < 0. \end{cases}$$

Similar expansions can be found for $W_{\kappa,\mu}(z)$ when $\mu = 0, 1/2, 1, \cdots$ as well.

Singularity Perturbation Analysis. We will now explain our method of proof of our limit laws from Section 1. The method is based on singularity perturbation analysis, a term coined by Flajolet and Lafforge [14]. The idea is to directly work with the moment-generating function of X_n which by Cauchy's integral formula is obtained from $X(y,z)$ by

$$\mathbb{E}\left(e^{yX_n}\right) = \frac{1}{2\pi i}\int_\gamma \frac{X(y,z)}{z^{n+1}}\mathrm{d}z. \tag{13}$$

Here, y is considered to be a parameters for which we assume that $|y| < \eta$ with $\eta > 0$ suitable small.

In order to use (13), one has to choose a suitable contour γ and to study the singularity structure of $X(y,z)$. Due to the above explicit expression for $X(y,z)$, we see that the singularities are either the branch point singularities (with moving branch-cut) of the Whittaker functions or are poles arising from the zeros of $V(y,z)$. With a change of variable, we can consider

$$\begin{aligned}\tilde{V}(y,z) =& M_{-(1-p)e^{y}/2,\sqrt{1-4p(1-p)e^{y}}/2}(2(1-p)(z-1)) \\ &+ c(y)W_{-(1-p)e^{y}/2,\sqrt{1-4p(1-p)e^{y}}/2}(2(1-p)(z-1)). \end{aligned} \tag{14}$$

Now, the branch cut is fixed at $[1,\infty)$. As for the zeros of this function, we will prove that there are two cases:

- Case I: $p = 0$. Here, we will show that for $|z| < 1+\delta$ with a suitable δ, we have exactly one zero $z_0(y)$ of $\tilde{V}(y,z)$. Moreover, this zero has the property that it converges to the branch point singularity as y tends to 0.
- Case II: $p > 0$. Here, we will show that for $|z| < 1+\delta$ with a suitable δ, we have no zeros of $\tilde{V}(y,z)$.

The second case is more in line with other instances to which singularity perturbation analysis was applied; see [14] and Chapter X of [15]. In this case, γ will be deformed in the standard contour from singularity analysis; see the right contour in Figure 3. The asymptotic evaluation of (13) is then immediate and the main term comes from the part of the contour close to the branch point singularity.

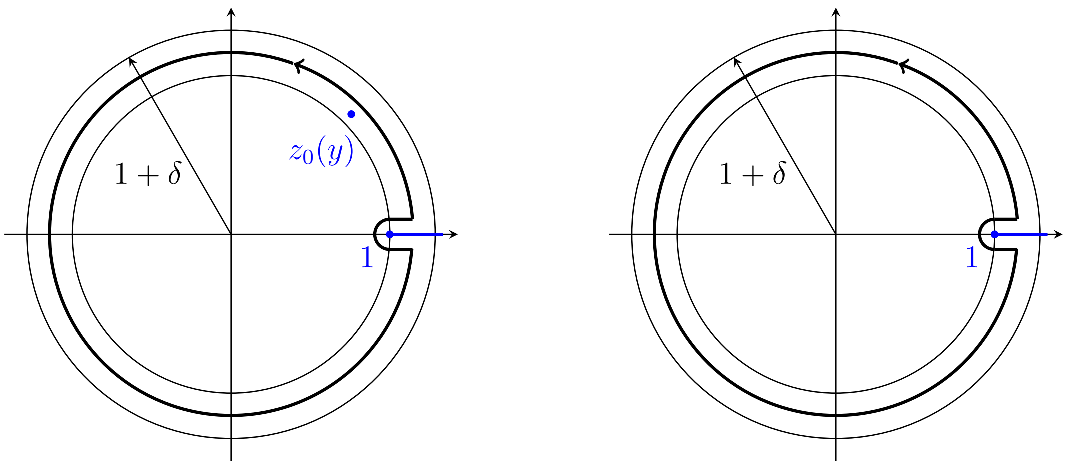

Figure 3: The integration contour and singularities (in blue) in the two cases. The only (but crucial) difference is the additional polar singularity in Case I.

The first case is more involved, in particular, due to the fact that the polar singularity (arising from the zero of $\tilde{V}(y,z)$) coalesces with the branch point singularity as y tends to 0. Note that a somehow similar situation was encountered in a recent study of Drmota et al. [10]. In fact, our approach in Case I will resemble the one of [10]. More precisely, we will again deform the contour into the same type of contour as in Case II; see the left contour in Figure 3. This will lead to a contribution coming from the polar singularity by a straightforward application of the the residue theorem. Then, in contrast to Case II, we will show that the contribution of the branch point singularity is negligible. From this, the unusual central limit theorem of the neutral model will follow.

Remark 6. Analytically, the unusual normalization in Theorem 1 arises from the two coalescing singularities. If, for instance, one would only have a polar singularity, then a central limit theorem with the usual normalization would hold; see Flajolet et al. [13].

4 Limit Laws

In this section, we will prove Theorem 1-4. For the proof, we will use singularity perturbation analysis and the properties of the Whittaker functions from the previous section. Moreover, in the cases $0 < p < 1/2$ and $1/2 < p \leq 1$, an alternative proof will be given with the method of moments.

We start with the most interesting case of the neutral model.

Proof of Theorem 1: $p = 0$. We first collect some properties of $\tilde{V}(y, z)$ and $\tilde{V}'(y, z)$ (for the definition see (14)).

Lemma 8. *Let $|y| < \eta$ and*

$$\tilde{\Delta} = \{z \in \mathbb{C} \ : \ |z| < 1 + \delta, \ \arg(z - 1) \neq 0\},$$

where $\eta, \delta > 0$. Then, $\tilde{V}(y, z)$ and $\tilde{V}'(y, z)$ are analytic in $\tilde{\Delta}$ and satisfy

$$\tilde{V}(y, z) = 2(z - 1) + 2ay + 4ay(z - 1)\log(z - 1) + \mathcal{O}\left(\max\{y^2, y(z - 1), (z - 1)^2\}\right), \tag{15}$$

$$\tilde{V}'(y, z) = 2 + 4ay\log(z - 1) + \mathcal{O}\left(\max\{y, z - 1\}\right), \tag{16}$$

where $a = (1 - e^{-2})/4$.

Proof. This follows from the properties of the Whittaker function from Section 3. ▮

Next, we need the following lemma which was already announced in the previous section.

Lemma 9. *For η, δ sufficiently small, $\tilde{V}(y, z)$ as a function of z has only one (simple) zero $z_0(y)$ in $\tilde{\Delta}$. Moreover, we have have, as $y \to 0$,*

$$z_0(y) = 1 - ay + 2a^2y^2\log y + \mathcal{O}(y^2).$$

Proof. First note that $\tilde{V}(0, z) = 2(z - 1)e^{z-1}$ which is an entire function with only one (simple) zero at $z = 1$. Next, $\tilde{V}(y, z)$ is analytic in both y and z in $\tilde{\Delta}$ and thus its zeros vary continuously with y. Also, note that because of (15) of the above lemma, there is no zero in a sufficiently small neighborhood of $z = 1$ for y sufficiently small (the limits as z tends to the branch-cut in the neighborhood are never equal to zero as well). Thus, for η, δ sufficiently small, we exactly have one zero in $\tilde{\Delta}$ which in addition must move to 1 as y tends to 0. This proves the first claim.

As for the proof of the second claim, we use bootstrapping. We already know that $z_0(y) = 1 + o(y)$. Plugging this into (15), we obtain that, as $y \to 0$,

$$z_0(y) = 1 - ay + o(y).$$

Using another bootstrapping step, this can be refined to

$$z_0(y) = 1 - ay + 2a^2y^2\log y + o(y^2\log y).$$

Yet another bootstrapping step gives the following refined error bound

$$z_0(y) = 1 - ay + 2a^2y^2\log y + \mathcal{O}(y^2).$$

This is the second claim. ▮

Now, we can show the key lemma for the proof of Theorem 1.

Lemma 10. *Let $y = it/(2a\sqrt{n \log n})$. Then,*

$$\mathbb{E}\left(e^{yX_n}\right) = z_0(y)^{-n} + \mathcal{O}\left(\frac{1}{\log n}\right). \tag{17}$$

Proof. For the proof, we use Cauchy's integral formula

$$\begin{aligned}
\mathbb{E}\left(e^{yX_n}\right) &= [z^n]X(y,z) = -[z^{n-1}]\frac{V'(y,z)}{V(y,z)} \\
&= -\frac{1}{2\pi i}\int_{\tilde{\gamma}} \frac{V'(y,z)}{V(y,z)}\frac{\mathrm{d}z}{z^n} \\
&= -\frac{1}{2\pi i}\int_{\gamma} \frac{\tilde{V}'(y,w)}{\tilde{V}(y,w)}\frac{\mathrm{d}\omega}{(e^{-y/2}(w-1)+1)^n},
\end{aligned}$$

where $\tilde{\gamma}$ is a small positively oriented circle centered at the origin and the last step follows from the change of variables $e^{y/2}(z-1) = w-1$. We now deform the countour γ into a contour γ' which is given by $\gamma' = \gamma_1' \cup \gamma_2'$ with

$$\gamma_1' = \{w = 1 + v/n \ : \ v \in \mathcal{H}_n\},$$

where $\mathcal{H}_n$ denotes the major part of the Hankel contour with

$$\mathcal{H}_n = \{v \in \mathbb{C} \ : \ |v| = 1, \Re(v) \le 0\} \cup \{v \in \mathbb{C} \ : \ 0 \le \Re(v) \le \sqrt{(1+\delta')^2 n^2 - 1} - n, \Im(v) = \pm 1\}$$

and γ_2' completes the contour with an almost circle of radius $1+\delta'$ with $0 < \delta' < \delta$; see Figure 3. Note that the above integral then becomes

$$\mathbb{E}\left(e^{yX_n}\right) = (e^{-y/2}(z_0(y)-1)+1)^{-n} - \frac{1}{2\pi i}\int_{\gamma'} \frac{\tilde{V}'(y,w)}{\tilde{V}(y,w)}\frac{\mathrm{d}\omega}{(e^{-y/2}(w-1)+1)^n}$$

since by the residue theorem, we have to add the residue

$$\operatorname{Res}\left(\frac{\tilde{V}'(y,w)}{\tilde{V}(y,w)}(e^{-y/2}(w-1)+1)^{-n}, w = z_0(y)\right) = (e^{-y/2}(z_0(y)-1)+1)^{-n}$$

In order to derive (17) from this, we first note that from $z_0(y) = 1 + \mathcal{O}(1/\sqrt{n \log n})$, we obtain that

$$(e^{-y/2}(z_0(y)-1)+1)^{-n} = z_0(y)^{-n}\left(1 + \mathcal{O}\left(\frac{1}{n \log n}\right)\right)^{-n} = z_0(y)^{-n} + \mathcal{O}\left(\frac{1}{\log n}\right).$$

Next for the integral, note that by (15) and (16) and again $z_0(y) = 1+\mathcal{O}(1/\sqrt{n \log n})$, we have for $w \in \gamma_1'$,

$$\frac{\tilde{V}'(y,w)}{\tilde{V}(y,w)} = \mathcal{O}\left(\frac{n}{\log^2 n}\right).$$

Moreover, for $w \in \gamma_1'$

$$|e^{-y/2}(w-1)+1|^{-n} \le \left(1 + \frac{\Re(e^{-y/2}v)}{n}\right)^{-n} \le e^{-\Re(e^{-y/2}v)} = \mathcal{O}\left(e^{-\epsilon\Re(v)}\right) \tag{18}$$

for a suitable $\epsilon > 0$. Hence, we obtain that for $w \in \gamma_1'$,

$$\frac{\tilde{V}'(y,w)}{\tilde{V}(y,w)} \cdot \frac{1}{(e^{-y/2}(w-1)+1)^n} = \mathcal{O}\left(\frac{n}{\log^2 n} e^{-\epsilon \Re(v)}\right).$$

Consequently,

$$-\frac{1}{2\pi i}\int_{\gamma_1'} \frac{\tilde{V}'(y,w)}{\tilde{V}(y,w)} \frac{\mathrm{d}\omega}{(e^{-y/2}(w-1)+1)^n} = -\frac{1}{2\pi i}\int_{\mathcal{H}_n} \mathcal{O}\left(\frac{n}{\log^2 n} e^{-\epsilon \Re(v)}\right)\frac{\mathrm{d}v}{n} = \mathcal{O}\left(\frac{1}{\log^2 n}\right).$$

Finally, suppose that $|w| = 1+\delta'$. First, from the analyticity of $\tilde{V}(y,z)$ and $\tilde{V}'(y,z)$, we obtain that

$$\frac{\tilde{V}'(y,z)}{\tilde{V}(y,z)} = \mathcal{O}(1).$$

Moreover,

$$|e^{-y/2}(w-1)+1| \geq |w| + \mathcal{O}\left(\frac{1}{\sqrt{n}\log n}\right) \geq 1+\delta''$$

for n large enough with $0 < \delta'' < \delta'$. Thus,

$$-\frac{1}{2\pi i}\int_{\gamma_2'} \frac{\tilde{V}'(y,w)}{\tilde{V}(y,w)} \frac{\mathrm{d}\omega}{(e^{-y/2}(w-1)+1)^n} = \mathcal{O}\left((1+\delta'')^{-n}\right).$$

Putting everything gives

$$\begin{aligned}\mathbb{E}\left(e^{yX_n}\right) &= z_0(y)^{-n} + \mathcal{O}\left(\frac{1}{\log n}\right) + \mathcal{O}\left(\frac{1}{\log^2 n}\right) + \mathcal{O}\left((1+\delta'')^{-n}\right) \\ &= z_0(y)^{-n} + \mathcal{O}\left(\frac{1}{\log n}\right)\end{aligned}$$

which is the claimed result. ∎

The proof of Theorem 1 follows now from the last lemma.

Proof of Theorem 1. As in Lemma 10 set $y = it/(2a\sqrt{n}\log n)$. Then, by the expansion of $z_0(y)$ from Lemma 9, we obtain that

$$z_0(y) = 1 - \frac{it}{2\sqrt{n}\log n} + \frac{t^2}{4n} + \mathcal{O}\left(\frac{\log\log n}{n\log n}\right).$$

Inserting this into the result from Lemma 10 yields

$$\mathbb{E}\left(e^{yX_n}\right) = \exp\left(\frac{it\sqrt{n}}{2\sqrt{\log n}} - \frac{t^2}{4}\right)\left(1 + \frac{\log\log n}{\log n}\right)$$

and by rearranging

$$\mathbb{E}\left(e^{it(X_n - an)/(2a\sqrt{n}\log n)}\right) = \exp\left(-\frac{t^2}{4}\right)\left(1 + \frac{\log\log n}{\log n}\right).$$

Since $\exp(-t^2/4)$ is the characteristic function of a normal distribution with mean 0 and variance $1/2$, the claimed central limit theorem follows from this by Lévy's continuity theorem. ∎

Proof of Theorem 2: $0 < p < 1/2$**.** The proof of weak convergence will follow with a similar line of reasoning as in the last paragraph. Thus, we again start by giving expansions for $\tilde{V}(z, y)$ and $\tilde{V}'(z, y)$.

Lemma 11. *Let* $|y| < \eta$ *and*

$$\tilde{\Delta} = \{z \in \mathbb{C} \ : \ |z| < 1 + \delta, \arg(1 - z) \neq \pi\},$$

where $\eta, \delta > 0$*. Then,* $\tilde{V}(y, z)$ *and* $\tilde{V}'(y, z)$ *are analytic in* $\tilde{\Delta}$ *and satisfy*

$$\begin{aligned}\tilde{V}(y, z) &= 2^{1-p}(1-p)^{1-p}(z-1)^{1-p} - \frac{2^{p-1}(1-p)^{p+1}}{(1-2p)^2}m(p)y(z-1)^p \\ &\quad + \mathcal{O}\left(\max\{y^2(z-1)^p, y(z-1)^{1-p}, (z-1)^{2-p}\}\right), \qquad (19)\\ \tilde{V}'(y, z) &= 2^{1-p}(1-p)^{2-p}(z-1)^{-p} - \frac{p2^{p-1}(1-p)^{p+1}}{(1-2p)^2}m(p)y(z-1)^{p-1} \\ &\quad + \mathcal{O}\left(\max\{y^2(z-1)^{p-1}, y(z-1)^{-p}, (z-1)^{1-p}\}\right). \qquad (20)\end{aligned}$$

Proof. This follows from the properties of Whittaker functions from Section 3. ▮

Next, we need to study zeros of $\tilde{V}(y, z)$. In contrast to Lemma 9, in the current case, we have no zeros.

Lemma 12. *For* η, δ *sufficiently small,* $\tilde{V}(y, z)$ *as a function of* z *has no zeros in* $\tilde{\Delta}$*.*

Proof. First note that

$$\tilde{V}(0, z) = 2^{1-p}(1-p)^{1-p}(z-1)^{1-p}e^{-(1-p)(z-1)}$$

which has no zero in $\tilde{\Delta}$ and only tends to 0 on the branch-cut when z tends to 1. The latter property holds for $\tilde{V}(y, z)$ for η and δ small enough as well as can be easily seen from (19). Thus, due to the analyticity of $\tilde{V}(y, z)$, all its zeros have to escape to infinity as y tends to zero. Consequently, for η sufficiently small, $\tilde{V}(y, z)$ has no zero in $\tilde{\Delta}$. ▮

The main lemma in this context is the following one.

Lemma 13. *Let* $y = it/n^{1-2p}$*. Then,*

$$\mathbb{E}\left(e^{yX_n}\right) = \frac{1}{2\pi i}\int_{\mathcal{H}} \Phi(y, v)e^{-v}\mathrm{d}v + \mathcal{O}\left(\frac{\log^2 n}{n^{1-2p}}\right),$$

where $\mathcal{H}$ *is the Hankel contour and* $\Phi(y, v)$ *was defined in Theorem 2.*

Proof. The proof is similar to the proof of Lemma 10 with the crucial difference that now the main contribution will come from the branch-point singularity (since there is no polar singularity). The starting point is again Cauchy's integral formula which as in Lemma 10 can be rewritten to

$$\mathbb{E}\left(e^{yX_n}\right) = -\frac{1}{2(1-p)\pi i}\int_{\gamma} \frac{\tilde{V}'(y, w)}{\tilde{V}(y, w)} \frac{\mathrm{d}w}{(e^{-y/2}(w-1)+1)^n}.$$

We again deform the contour γ into a contour γ' which this time is $\gamma' = \gamma_1' \cup \gamma_2' \cup \gamma_3'$, where

$$\gamma_i' = \{w = 1 + v/n \ : \ v \in \mathcal{H}_n^{(i)}\}, \qquad (i = 1, 2)$$

with $\mathcal{H}_n^{(i)}, i = 1, 2$, given by

$$\begin{aligned}\mathcal{H}_n^{(1)} &= \{v \in \mathbb{C} \ : \ |v| = 1, \Re(v) \leq 0\} \cup \{v \in \mathbb{C} \ : \ 0 \leq \Re(v) \leq \log^2 n, \Im(v) = \pm 1\},\\ \mathcal{H}_n^{(2)} &= \{v \in \mathbb{C} \ : \ \log^2 n < \Re(v) \leq \sqrt{(1+\delta')^2 n^2 - 1} - n, \Im(v) = \pm 1\}\end{aligned}$$

and γ_3' completes the contour with an almost circle of radius $1+\delta'$ with $\delta' < \delta$. The difference to the contour from Lemma 10 is that $\mathcal{H}_n$ there is split into the two parts $\mathcal{H}_n^{(1)}$ and $\mathcal{H}_n^{(2)}$. Moreover, note that now, deforming the contour in the integral above will leave the value of the integral unchanged.

We proceed by treating the three integrals corresponding to the above three parts of the contour. First, for γ_1' note that from (19) and (20), we obtain that

$$\frac{\tilde{V}'(y,w)}{\tilde{V}(y,w)} = n(1-p)\Phi(it,v)\left(1+\mathcal{O}\left(\frac{\log^{2-2p} n}{n^{1-2p}}\right)\right).$$

Moreover, we have that

$$(e^{-y/2}(w-1)+1)^{-n} = e^{-v}\left(1+\mathcal{O}\left(\frac{\log^2 n}{n^{1-2p}}\right)\right).$$

Thus,

$$\begin{aligned}
-\frac{1}{2(1-p)\pi i}&\int_{\gamma_1'}\frac{\tilde{V}'(y,z)}{\tilde{V}(y,z)}\frac{\mathrm{d}w}{(e^{-y/2}(w-1)+1)^n}\\
&= -\frac{1}{2\pi i}\int_{\mathcal{H}_n^{(1)}} n\Phi(it,v)e^{-v}\left(1+\mathcal{O}\left(\frac{\log^{2-2p} n}{n^{1-2p}}\right)\right)e^{-v}\left(1+\mathcal{O}\left(\frac{\log^2 n}{n^{1-2p}}\right)\right)\frac{\mathrm{d}v}{n}\\
&= -\frac{1}{2\pi i}\int_{\mathcal{H}_n^{(1)}}\Phi(it,v)e^{-v}\mathrm{d}v + \mathcal{O}\left(\frac{\log^2 n}{n^{1-2p}}\right)\\
&= \frac{1}{2\pi i}\int_{\mathcal{H}}\Phi(it,v)e^{-v}\mathrm{d}v + \mathcal{O}\left(\frac{\log^2 n}{n^{1-2p}}\right),
\end{aligned}$$

where the last step follows by attaching the tails of the Hankel contour (which introduces a negligible error) and changing the orientation of the contour.

Next, we consider the contribution of the integral over γ_2'. Here, we have

$$\frac{\tilde{V}'(y,w)}{\tilde{V}(y,w)} = \mathcal{O}(1).$$

Moreover, by using (18) which holds in the current situation as well, we obtain that

$$-\frac{1}{2(1-p)\pi i}\int_{\gamma_2'}\frac{\tilde{V}'(y,z)}{\tilde{V}(y,z)}\frac{\mathrm{d}w}{(e^{-y/2}(w-1)+1)^n} = -\frac{1}{2\pi i}\int_{\mathcal{H}_n^{(2)}}\mathcal{O}\left(e^{-\epsilon\Re(v)}\right)\frac{\mathrm{d}v}{n} = \mathcal{O}\left(\frac{1}{n}\right).$$

Finally, the integral over γ_3' is exactly treated as in Lemma 10 and it contributes only an exponential decreasing error term. Collecting these three parts yields the claimed result. ∎

Now, we can complete the proof of Theorem 2.

Proof of Theorem 2. Weak convergence follows from the previous lemma.

In order to show that also moments converge, we work with the moments from Proposition 2 and use the method of moments. Accordingly, the only thing which one has to verify is that there is unique random variable whose moment sequence is given by $d_k/\Gamma(k(1-2p)+1)$. For this purpose, it suffices to show that

$$\sum_{k\geq 1}\frac{d_k}{\Gamma(k(1-2p)+1)}z^k$$

has a positive radius of convergence. This clearly follows from the estimate

$$d_k \le A^k k! k^{k(1-2p)} \tag{21}$$

for a sufficiently large A.

We will prove (21) by induction. First, by suitable choosing A, it is clear that we can assume that the estimate holds for all small k. Now, assume that it holds for all $k' < k$. In order to prove it for k, we insert the induction hypothesis into the recurrence for d_k from Proposition 2. This gives

$$\begin{aligned} d_k &\le A^k k! \frac{2(1-p)}{k(k-1)(1-2p)} \sum_{j=1}^{k-1} j(k-j)^{(k-j)(1-2p)} j^{j(1-2p)} \\ &\le A^k k! \frac{2(1-p)}{k(1-2p)} \sum_{j=1}^{k-1} \left((k-j)^{k-j} j^j\right)^{1-2p}. \end{aligned}$$

Now, note that $(k-j)^{k-j} j^j$ is decreasing for $0 < j \le j/2$. Choose j_0 such that $j_0 > 1/(1-2p)$. Then,

$$d_k \le A^k k! \frac{2(1-p)}{k(1-2p)} \left(2j_0 k^{(k-1)(1-2p)} + k^{1+(k-j_0)(1-2p)} j_0^{j_0(1-2p)}\right) \le A^k k! k^{k(1-2p)},$$

where the last inequality holds for k large enough. This concludes the induction step.

Finally, since we know already that X_n/n^{1-2p} weakly converges to X, we must have that $\mathbb{E}(X^k) = d_k/\Gamma(k(1-2p)+1)$. Thus,

$$\mathbb{E}\left(e^{yX}\right) = \frac{1}{2\pi i} \int_{\mathcal{H}} \Phi(y,v) e^{-v} \mathrm{d}v.$$

This concludes the proof of Theorem 2. ∎

We conclude this paragraph with the proof of Corollary 1.

Proof of Corollary 1. It is sufficient to show that

$$\int_0^\infty f(x) e^{yx}\,\mathrm{d}x = \frac{1}{2\pi i} \int_{\mathcal{H}} \left(\Phi(y,v) - \frac{p}{v(1-p)}\right) e^{-v} \mathrm{d}v$$

for every fixed $y \in \mathbb{C}$. For this purpose, we use the series representation (3) and Hankel's representation of the reciprocal of the Gamma function

$$\frac{1}{\Gamma(2(k+1)p-k)} = -\frac{1}{2\pi i} \int_{\mathcal{H}} (-v)^{-2(k+1)p+k} e^{-v} \mathrm{d}v.$$

Next, we replace the Hankel contour $\mathcal{H}$ by the contour $\mathcal{H}'$ that starts in the upper half plane at $+e^{i\varphi}\infty$, winds around 0 counterclockwise before it tends to $+e^{-i\varphi}\infty$ in the lower half plane, where $0 < \varphi < \pi/2$ is chosen such that $(\pi-\varphi)(1-2p) < \pi/2$. In particular, we can choose $\mathcal{H}'$ in a way that $\Re(\delta(-v)^{1-2p}+y) < 0$ for all $v \in \mathcal{H}'$; note that $\delta = \delta(p) < 0$.

Hence, after interchanging the integral and the series and by evaluating the exponential series

$$\sum_{k\ge 0} \frac{(\delta(-v)^{1-2p} x)^k}{k!} = e^{\delta(-v)^{1-2p} x}$$

we can compute the (inner) integral

$$\int_0^\infty e^{(\delta(-v)^{1-2p}+y)x} = \frac{-1}{\delta(-v)^{1-2p}+y}$$

and finally get

$$\int_0^\infty f(x)e^{yx}\,\mathrm{d}x = \frac{1}{2\pi i}\int_{\mathcal{H}'}\left(\Phi(y,v) - \frac{p}{v(1-p)}\right)e^{-v}\mathrm{d}v.$$

It is clear that $\mathcal{H}'$ can be (again) replaced by $\mathcal{H}$ and so the result follows. ▮

Proof of Theorem 3 and Theorem 4: $1/2 \le p < 1$. We will consider here the proofs of Theorem 3 and Theorem 4 which will merged. For the weak convergence, we will proceed as in the previous paragraph. The following two lemmas can be proved with the same method as before.

Lemma 14. *Let $|t| < \eta$ and*

$$\tilde{\Delta} = \{z \in \mathbb{C} \,:\, |z| < 1+\delta, \arg(1-z) \neq \pi\},$$

where $\eta, \delta > 0$. Then, $\tilde{V}(it,z)$ and $\tilde{V}(it,z)$ are analytic in $\tilde{\Delta}$ and satisfy

$$\tilde{V}(it,z) = d(t)(z-1)^{\left(1-\sqrt{1-4p(1-p)e^{it}}\right)/2} + \mathcal{O}\left((z-1)^{\left(1+\sqrt{1-4p(1-p)e^{it}}\right)/2}\right), \tag{22}$$

$$\begin{aligned}\tilde{V}'(it,z) = {} & d(t)\left(\frac{1-\sqrt{1-4p(1-p)e^{it}}}{2}\right)(z-1)^{\left(-1-\sqrt{1-4p(1-p)e^{it}}\right)/2} \\ & + \mathcal{O}\left((z-1)^{\left(-1+\sqrt{1-4p(1-p)e^{it}}\right)/2}\right),\end{aligned} \tag{23}$$

where

$$d(t) = \frac{\Gamma(\sqrt{1-4p(1-p)e^{it}}/2)}{\Gamma(1/2+(1-p)e^{it/2}+\sqrt{1-4p(1-p)e^{it}}/2)}(2(1-p))^{\left(1+\sqrt{1-4p(1-p)e^{it}}\right)/2}c(it).$$

Lemma 15. *For η, δ sufficiently small, $\tilde{V}(it,z)$ as a function of z has no zeros in $\tilde{\Delta}$.*

From these two lemmas, we prove the following result.

Lemma 16. *Let $|t| < \eta$ with η sufficiently small. Then,*

$$\mathbb{E}\left(e^{itX_n}\right) \longrightarrow \frac{1-\sqrt{1-4p(1-p)e^{it}}}{2(1-p)}.$$

Proof. Obviously, we can assume that $|t| > 0$. The proof is then similar to the one of Lemma 13. We only highlight differences. First, as in the proof of Lemma 13, we obtain that

$$\mathbb{E}\left(e^{itX_n}\right) = -\frac{1}{2(p-1)\pi i}\int_\gamma \frac{\tilde{V}'(it,w)}{\tilde{V}(it,w)}\frac{\mathrm{d}w}{(e^{-it/2}(w-1)+1)^n}.$$

Then, we again deform the contour to $\gamma' = \gamma_1' \cup \gamma_2' \cup \gamma_3'$.

The treatment of the integral over γ_2' and γ_3' is completely the same as in the proof of Lemma 13. Thus, we only have to concentrate on γ_1'. Here, we have from (22) and (23),

$$\frac{\tilde{V}'(it,w)}{\tilde{V}(it,w)} = \frac{1-\sqrt{1-4p(1-p)e^{it}}}{2}nv^{-1}\left(1+\mathcal{O}\left(\left(\frac{\log^2 n}{n}\right)^{\Re(\sqrt{1-4p(1-p)e^{it}})}\right)\right).$$

Note that the above real part is positive for $|t|$ small (even in the boundary case $p = 1/2$). Moreover, we have

$$(e^{-it/2}(w-1)+1)^{-n} = e^{-e^{-it/2}v}\left(1+\mathcal{O}\left(\frac{\log^2 n}{n^2}\right)\right).$$

Plugging into the above integral yields

$$-\frac{1}{2(p-1)\pi i}\int_{\gamma_1}\frac{\tilde{V}'(it,w)}{\tilde{V}(it,w)}\frac{\mathrm{d}w}{(e^{-it/2}(w-1)+1)^n}$$
$$= \frac{1-\sqrt{1-4p(1-p)e^{it}}}{2(p-1)}\left(-\frac{1}{2\pi i}\int_{\mathcal{H}_n^{(1)}} v^{-1}e^{-e^{-it/2}v}\mathrm{d}v\right)+o(1).$$

The last integral can be reduced to

$$-\frac{1}{2\pi i}\int_{\mathcal{H}_n^{(1)}} v^{-1}e^{-e^{-it/2}v}\mathrm{d}v = \frac{1}{2\pi i}\int_{\mathcal{H}} v^{-1}e^{-v}\mathrm{d}v + o(1) = 1+o(1),$$

where the last step follows from Hankel's integral representation of $1/\Gamma(z)$. Thus, the part of the integral over γ_1' gives the main contribution and the result follows. ∎

We can now finish the proof of Theorem 3 and Theorem 4.

Proof of Theorem 3 and Theorem 4. The weak convergence part follows from the last lemma.

Next, we prove that when $1/2 < p < 1$, then also all moments converge. To this end, as in the proof of Theorem 2, we only need to show that with the e_k's from Proposition 4, the following series

$$E(z) = \sum_{k\geq 1} e_k \frac{z^k}{k!}$$

has a positive radius of convergence. In fact, using the recurrence for e_k, we can directly show that

$$E(z) = \frac{\sqrt{1-4p(1-p)e^z}}{2} - 1$$

as must be the case. To prove this, note that the recurrence for e_k implies that

$$(2p-1)E'(z) - p = 2(1-p)E(z)E'(z) + p(e^z-1)$$

with $E(0) = 0$. Integrating gives

$$(2p-1)E(z) - pz = (1-p)E(z) + p(e^z - z - 1).$$

Thus,

$$E(z) = \frac{2p-1-\sqrt{(2p-1)^2-4p(1-p)(e^{z-1}-1)}}{2(1-p)} = \frac{1-\sqrt{1-4p(1-p)e^z}}{2(1-p)} - 1.$$

This proves the claimed result. ∎

We conclude by pointing out that from the explicit expression for the characteristic function of X, one can find the probability mass function by simple Taylor series expansion.

Corollary 2 ($1/2 \leq p \leq 1$)**.** *The limiting distribution X is a discrete distribution on $\{1, 2, \ldots\}$ with*

$$P(X = k) = \frac{p^k(1-p)^{k-1}}{2(2k-1)}\binom{2k}{k}.$$

5 Conclusion

In this paper, we gave a detailed analysis of the extra clustering model which was recently introduced by Durand et al. [7] because of two reasons: (i) to model the group formation process of social animals and (ii) to test whether genetic relatedness is the main driving force behind the group formation process. Our analysis extends the previous analysis of [8] which was concerned with asymptotic expansions of the mean of the number of groups formed by the animals. We derived all higher moments and completely classified the limiting distribution of the number of groups for all values of p.

Our results show that the limiting distribution for the neutral model ($p = 0$) is a continuous distribution, the limiting distribution when $0 < p < 1/2$ is a mixture of discrete and continuous distributions and finally for $1/2 \leq p \leq 1$ the limiting distribution is a discrete random variable. This transition from continuous to discrete is in fact expected since the extra clustering model is getting less random as p increases (because animals are more likely to form one huge group).

From a mathematical point of view, our results contain two surprises. First, not in all cases, the limit law can be obtained by the method of moments. In fact, we have seen two cases, namely, $p = 0$ and $p = 1/2$, where we have weak convergence but moments do not converge. The case of the neutral model is in particular surprising because the underlying sequence of random variables satisfies a divide-and-conquer recurrence of a type which often appears in computer science and for which in many previous studies an application of the method of moments lead to the limit law. The second surprise is the curious limit law for the neutral model. In fact, our proof does not give a lot of insight of as to why this surprising result holds. A better explanation was given in a recent paper of Janson [17]. However, many things about this result are still shrouded in mystery, in particular, whether such a surprising result also holds for other classes of random trees such as random m-ary search trees (in this work, we considered trees which are equivalent to random binary search trees; for a definition of this family of trees as well as random m-ary search trees see Mahmoud [20]). We hope to come back to this question in a future work.

Acknowledgements

An extended abstract [9] and talk containing preliminary materials of this paper were presented at the 25th International Meeting on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms which took place in Paris from June 16 to June 20, 2014. The authors thank the reviewers of this abstract and the participants of the meeting for valuable suggestions and a lot of helpful input. The first author acknowledges partial support by the Austrian Science Foundation, SFB F50 "Algorithmic and Enumerative Combinatorics". The second author was partially supported by the Ministry of Science and Technology, Taiwan under the grant MOST-103-2115-M-009-007-MY2.

References

[1] R. Beals and R. Wong. *Special Functions: A Graduate Text*, Cambridge Studies in Advanced Mathematics, 126, Cambridge University Press, Cambridge, 2010.

[2] P. Billingsley. *Probability and Measure*, third edition, Wiley Series in Probability and Mathematical Statistics, A Wiley-Interscience Publication, John Wiley & Sons, Inc., New York, 1995.

[3] M. G. B. Blum and O. François (2005). Minimal clade size and external branch length under the neutral coalescent, *Adv. in Appl. Probab.*, **37:3**, 647–662.

[4] M. Blum, O. François, S. Janson (2006). The mean, variance and limiting distributions of two statistics sensitive to phylogenetic tree balance, *Ann. Appl. Probab.*, **16:4**, 2195–2214.

[5] H. Chang and M. Fuchs (2010). Limit theorems for patterns in phylogenetic trees, *J. Math. Biol.*, **60:4**, 481–512.

[6] H.-H. Chern, M. Fuchs, H.-K. Hwang (2007). Phase changes in random point quadtrees, *ACM Trans. Algorithms*, **3:3**, 51 pages.

[7] E. Durand, M. G. B. Blum and O. Francçis (2007). Prediction of group patterns in social mammals based on a coalescent model, *J. Theoret. Biol.*, **249:2**, 262–270.

[8] E. Durand and O. Francçis (2010). Probabilistic analysis of a genealogical model of animal group patterns, *J. Math. Biol.*, **60:3**, 451–468.

[9] M. Drmota, M. Fuchs, Y.-W. Lee (2014). Limit laws for the number of groups formed by social animals under the extra clustering model, In Proceedings of the 25th International Meeting on Probabilistic, Combinatorial and Asymptotic Methods for the Analysis of Algorithms, *Discrete Math. Theor. Comput. Sci. Proc.*, 73–85.

[10] M. Drmota, A. Iksanov, M. Möhle, U. Rösler (2009). A limiting distribution for the number of cuts needed to isolate the root of a random recursive tree, *Random Structures Algorithms*, **34:3**, 319–36.

[11] J. A. Fill, P. Flajolet, N. Kapur (2004). Singularity analysis, Hadamard products, and tree recurrences, *J. Comput. Appl. Math.*, **174:2**, 271–313.

[12] J. A. Fill and N. Kapur (2004). Limiting distributions for additive functionals on Catalan trees, *Theor. Comput. Sci.*, **326:1-3**, 69–102.

[13] P. Flajolet, X. Gourdon, C. Martinez (1997). Patterns in random binary search trees, *Random Structrues Algorithms*, **11:3**, 223–244.

[14] P. Flajolet and T. Lafforgue (1994). Search costs in quadtrees and singularity perturbation asymptotics, *Discrete Comput. Geom.*, **12:2**, 151–175.

[15] P. Flajolet and R. Sedgewick. *Analytic Combinatorics*, Cambridge University Press, Cambridge, 2009.

[16] H.-K. Hwang and R. Neininger (2002). Phase change of limit laws in the quicksort recurrences under varying toll functions, *SIAM J. Comput.*, **31:6**, 1687–1722.

[17] S. Janson (2014). Maximal clades in random binary search trees, preprint.

[18] S. Janson and G. Kersting (2011). On the total external length of the Kingman coalescent, *Electronic J. Probability*, **16**, 2203–2218.

[19] J. F. C. Kingman (1982). The coalescent, *Stochastic Process. Appl.*, **13:3**, 235–248.

[20] H. M. Mahmoud. *Evolution of Random Search Trees*. Wiley-Interscience, 1992.