HAL
archives-ouvertes.fr

Protected node profile of Tries

Mehri Javanian

▶ To cite this version:

Mehri Javanian. Protected node profile of Tries. Discrete Mathematics and Theoretical Computer Science, DMTCS, 2018, Vol. 20 no. 1. hal-01548175v6

HAL Id: hal-01548175
https://hal.archives-ouvertes.fr/hal-01548175v6

Submitted on 20 Mar 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Discrete Mathematics and Theoretical Computer Science DMTCS vol. **20**:1, 2018, #12

Protected node profile of tries

Mehri Javanian*

Department of Statistics, University of Zanjan, Iran

received 27th June 2017, revised 16th Feb. 2018, 1st Mar. 2018, accepted 6th Mar. 2018.

In a rooted tree, protected nodes are neither leaves nor parents of any leaves. They have some practical motivations, *e.g.*, in organizational schemes, security models and social-network models. In this paper, we introduce a new type of profile, namely, the protected node profile which counts the number of protected nodes with the same distance from the root in rooted trees. Here, we present the asymptotic expectations, variances, covariance and limiting bivariate distribution of protected node profile and non-protected internal node profile in random tries, an important data structure on words in computer science. Also we investigate the fraction of these expectations asymptotically. These results are derived by the methods of analytic combinatorics such as generating functions, Mellin transform, Poissonization and depoissonization, saddle point method and singularity analysis.

Keywords: tries, protected nodes, tree profiles, Poissonization, Mellin transform, recurrences, generating functions, singularity analysis, saddle point method

1 Introduction and Main Results

Tries (invented by de la Briandais) are fundamental tree data structures for re*trie*val of information. The information stored in a trie is a set of strings (see Knuth (1998) for more details). For simplicity, we only consider strings over a binary alphabet. The strings are stored in the leaves. More precisely, a trie is built on n infinite 0-1 strings as follows: if $n = 1$ then the only string is stored in the root as an external node; if $n > 1$, then the root is an internal node (empty node) and the strings with the first bit "0" ("1") are directed to the left (right) subtree; finally, the subtrees are constructed recursively by the same rules, but by removing the first bit of all strings (*cf.* Figure 1).

A random trie with n external nodes is a trie built over n infinite 0-1 strings (a trie of size n) generated by memoryless a source, that is, we assume each string is a Bernoulli i.i.d. sequence with success probability $0 < p < 1$ (the probability of occurring a "1"); we also use $q := 1 - p \leq p$. Random tries have been extensively studied; for more background, see Mahmoud (1992) or Park et al. (2009), and the references therein, for a thorough analysis of the profile (number of nodes at a given level) of tries.

By protected nodes, we mean the nodes with a distance of at least two to all the leaves. *E.g.*, Figure 1 shows the protected nodes in black color. Protected nodes were introduced by Cheon and Shapiro (2008) as a guide in various organizational schemes. For instance, if leaves represent customers it may

*javanian@znu.ac.ir

ISSN 1365–8050 © 2018 by the author(s) Distributed under a Creative Commons Attribution 4.0 International License

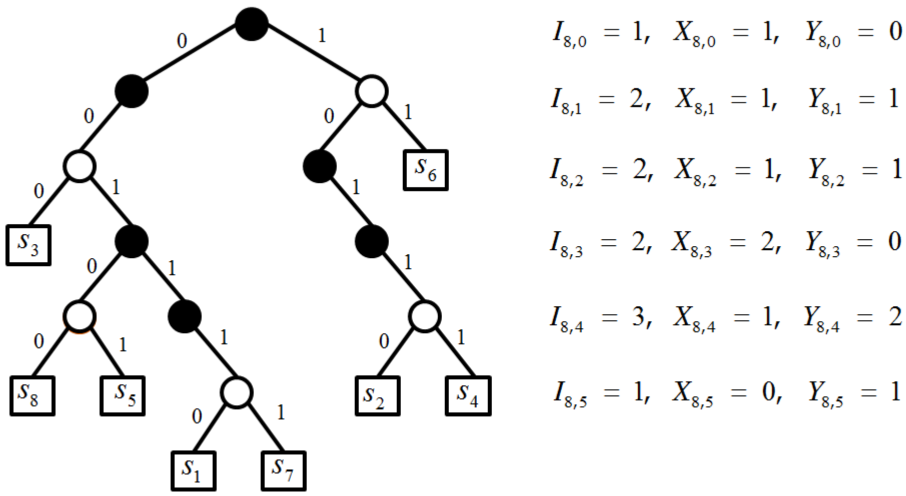

Fig. 1: A trie built on eight strings $s_1, \ldots, s_8$ (*i.e.*, $s_1 = 001110\ldots$, $s_2 = 10110\ldots$, $s_3 = 000\ldots$, *etc.*) with internal (circles), leaf (squares), protected (black circles), and non-protected (white circles) nodes, and its profiles.

be worthwhile for many of the points in the tree to be non-protected. However if the leaves represent lobbyists or computer hackers it may be a very good thing to have many points protected. In a security model with trie structure, a protected node may be taken to represent an entity that has at least two buffers between itself and a vulnerable point. Protected nodes have been investigated for many different random trees by many authors; see for instance, Du and Prodinger (2012), Devroye and Janson (2014), Fuchs et al. (2016) and the papers cited there.

For random tries, the mean and variance of the number of protected nodes have been obtained by Gaither et al. (2012) and Gaither and Ward (2013) where the above applications of this parameter such as security models with trie structures and social networks have been discussed. Moreover, Gaither and Ward (2013) announced a central limit theorem, which was conjectured in their paper. This conjecture has been confirmed by Fuchs et al. (2016) who proved (univariate and bivariate) central limit theorems for the number of protected nodes. Also, Fuchs et al. (2016) have shown that all previous results for tries can be derived by approaches of Hwang et al. (2012), Fuchs et al. (2014) and Fuchs and Lee (2014).

In the present paper, we are concerned with the protected node profile defined as the number of protected nodes with the same distance from the root in random tries. Throughout the paper, we write $I_{n,k}$, $X_{n,k}$ and $Y_{n,k}$ for, respectively, the number of internal nodes, the number of protected nodes and the number of non-protected internal nodes at level k in a trie of size n. Namely, $I_{n,k} = X_{n,k} + Y_{n,k}$ (*cf.* Figure 1). We also define $\gamma_{n,k} := \mathrm{Cov}(X_{n,k}, Y_{n,k})$ for the covariance of $X_{n,k}$ and $Y_{n,k}$; and

$$\mu_{n,k}^{[X]} := \mathbb{E}(X_{n,k}), \quad \mu_{n,k}^{[Y]} := \mathbb{E}(Y_{n,k}), \quad {\sigma_{n,k}^{[X]}}^2 := \mathbb{V}(X_{n,k}), \quad {\sigma_{n,k}^{[Y]}}^2 := \mathbb{V}(Y_{n,k}),$$

for the expectations and variances, respectively.

In order to state our findings and the results about $I_{n,k}$ in Park et al. (2009), we need the following notations. For a real number α with $(\log \frac{1}{q})^{-1} < \alpha < (\log \frac{1}{p})^{-1}$, let

$$\rho = \rho(\alpha) = \frac{1}{\log(p/q)} \log \frac{1 - \alpha \log(1/p)}{\alpha \log(1/q) - 1}. \tag{1}$$

Equivalently, α and ρ satisfy the equation

$$\alpha = \frac{p^{-\rho} + q^{-\rho}}{p^{-\rho}\log(1/p) + q^{-\rho}\log(1/q)}.$$

Furthermore, we set

$$\alpha_1 = \frac{1}{\log(1/q)}, \qquad \alpha_0 = \frac{2}{\log(1/p) + \log(1/q)},$$
$$\alpha_2 = \frac{p^2 + q^2}{p^2\log(1/p) + q^2\log(1/q)}, \qquad \beta(\rho) = \frac{p^{-\rho}q^{-\rho}\log(p/q)^2}{(p^{-\rho} + q^{-\rho})^2}.$$

The generic symbol ε represents a suitably small and positive constant whose value may vary from one occurence to another. The symbol $f(n) = \Theta(g(n))$ means that there are positive constants C and C' such that $C|g(n)| \le |f(n)| \le C'|g(n)|$.

We briefly recall the results for $\mathbb{E}(I_{n,k})$, $\mathbb{V}(I_{n,k})$ and limiting distribution of $I_{n,k}$ obtained by Park et al. (2009). For $p \neq q$, $\varepsilon > 0$, $\xi = o((\log n)^{\frac{1}{6}})$ and the normal distribution function $\Phi(x)$,

- If $1 \le k \le (\alpha_1 - \varepsilon)\log n$ then level k is almost full of internal nodes and $\mathbb{V}(I_{n,k}) \to 0$.
- If $(\alpha_1 + \varepsilon)\log n \le k \le (\alpha_0 - \varepsilon)\log n$ then $\rho_{n,k} := \rho(k/\log n) > 0$ and

$$\mathbb{E}(I_{n,k}) = 2^k - G^{[I]}\left(\rho_{n,k}, \log_{p/q} p^k n\right) \frac{(p^{-\rho_{n,k}} + q^{-\rho_{n,k}})^k n^{-\rho_{n,k}}}{\sqrt{2\pi\beta(\rho_{n,k})k}} \left(1 + \mathcal{O}(k^{-1/2})\right),$$

  where $G^{[I]}(\rho, x) = \sum_{j\in\mathbb{Z}} \Gamma(\rho + it_j)(\rho + it_j + 1)e^{-2\pi ijx}$ is a positive 1-periodic function and $t_j := 2\pi j/(\log p/q)$. The oscillating function $G^{[I]}(\rho, x)$ is consequence of an infinite number of saddle points appearing in the integrand of the associated Mellin transform of Poisson generating function of $\mathbb{E}(I_{n,k})$. This was first observed by Nicodème (2005). There is also an oscillating function in the formula of variance which is unbounded. The limiting distribution of $I_{n,k}$ is normal.
- If $k = \alpha_0(\log n + \xi\sqrt{\alpha_0\beta(0)\log n})$ then $\mathbb{E}(I_{n,k}) \sim 2^k\Phi(-\xi)$.
- If $(\alpha_0 + \varepsilon)\log n \le k \le (\alpha_2 - \varepsilon)\log n$ then $-2 < \rho_{n,k} := \rho(k/\log n) < 0$ and it is again the infinite number of saddle points that yield the dominant asymptotic approximation. Namely,

$$\mathbb{E}(I_{n,k}) \sim -G^{[I]}\left(\rho_{n,k}, \log_{p/q} p^k n\right) \frac{(p^{-\rho_{n,k}} + q^{-\rho_{n,k}})^k n^{-\rho_{n,k}}}{\sqrt{2\pi\beta(\rho_{n,k})k}},$$
$$(G^{[I]}(\rho, x) > 0, \text{ for } \rho > 0; \quad \text{and} \quad G^{[I]}(\rho, x) < 0, \text{ for } -2 < \rho < 0).$$

  Here, we recall there is a small mistake in the formula of $\mathbb{E}(I_{n,k})$ in Park et al. (2009): $G_3(\rho, x)(= G^{[I]}(\rho, x))$ should be multiplied by -1. The asymptotic variance has the same formula as the variance in the range $(\alpha_1 + \varepsilon)\log n \le k \le (\alpha_0 - \varepsilon)\log n$, and $I_{n,k}$ is again asymptotically normal.

- If $k = \alpha_2(\log n + \xi\sqrt{\alpha_2\beta(-2)\log n})$ then $\mathbb{E}(I_{n,k}) \sim \frac{1}{2}\Phi(\xi)n^2(p^2+q^2)^k$.
- If $k \geq (\alpha_2+\varepsilon)\log n$ then the oscillations appearing in the formula of $\mathbb{E}(I_{n,k})$ disappear since the behaviour of $\mathbb{E}(I_{n,k})$ is dominated by a polar singularity. We thus have $\mathbb{E}(I_{n,k}) \sim \frac{1}{2}n^2(p^2+q^2)^{k-1}$ and $\mathbb{V}(I_{n,k}) \sim \mathbb{E}(I_{n,k})$. Moreover, the limiting distribution of $I_{n,k}$ is normal, when $\mathbb{V}(I_{n,k}) \to \infty$, whereas the limiting distribution of $I_{n,k}$ is Poisson, when $\mathbb{V}(I_{n,k}) = \Theta(1)$.

In this paper, we consider the behavior of protected and non-protected internal node profiles in the range $(\alpha_1+\varepsilon)\log n \leq k \leq (\alpha_2-\varepsilon)\log n$. If $p=q$, then we have $\alpha_1=\alpha_0=\alpha_2$ and this range (the saddle point range) between α_1 and α_2 does not exist. Thus, we focus on the protected node and non-protected node profiles of asymmetric tries ($p \neq q$).

In Section 2, we first give a recurrence with respect to the joint probability generating function of $X_{n,k}$ and $Y_{n,k}$, and then the recurrences with respect to the Poissonized first two moments. The expected values, variances and covariance of our profiles are discussed in Section 2. We also show that the variances are of the same order as the expected values and therefore, are positive by the assertion of the following theorem.

Theorem 1.1 *Consider $\rho_{n,k} := \rho(k/\log n)$ and $t_j := 2\pi j/(\log p/q)$. For some $\varepsilon > 0$ and the functions*

$$\hat{g}^{[X]}(s) = 1 + s(p^{-s}+q^{-s}) - s(s+1)pq := -g^{[X]}(s),$$
$$g^{[Y]}(s) = p^{-s}+q^{-s} - 1 - (s+1)pq,$$

1. If $\alpha_1+\varepsilon \leq \frac{k}{\log n} \leq \alpha_0-\varepsilon$ then $\rho_{n,k} > 0$ and

$$\mu_{n,k}^{[X]} = 2^k - \hat{G}^{[X]}\left(\rho_{n,k}, \log_{p/q} p^k n\right)\frac{(p^{-\rho_{n,k}}+q^{-\rho_{n,k}})^k n^{-\rho_{n,k}}}{\sqrt{2\pi\beta(\rho_{n,k})k}}\left(1+\mathcal{O}(k^{-1/2})\right),$$

where $\hat{G}^{[X]}(\rho,x) = \sum_{j\in\mathbb{Z}}\Gamma(\rho+it_j)\hat{g}^{[X]}(\rho+it_j)e^{-2\pi ijx}$ is a positive 1-periodic function.

2. If $\alpha_0+\varepsilon \leq \frac{k}{\log n} \leq \alpha_2-\varepsilon$ then $-2 < \rho_{n,k} < 0$ and

$$\mu_{n,k}^{[X]} = G^{[X]}\left(\rho_{n,k}, \log_{p/q} p^k n\right)\frac{(p^{-\rho_{n,k}}+q^{-\rho_{n,k}})^k n^{-\rho_{n,k}}}{\sqrt{2\pi\beta(\rho_{n,k})k}}\left(1+\mathcal{O}(k^{-1/2})\right),$$

where $G^{[X]}(\rho,x) = \sum_{j\in\mathbb{Z}}\Gamma(\rho+it_j)g^{[X]}(\rho+it_j)e^{-2\pi ijx}$ is a positive 1-periodic function.

3. If $\alpha_1+\varepsilon \leq \frac{k}{\log n} \leq \alpha_2-\varepsilon$ then $\rho_{n,k} > -2$ and

$$\mu_{n,k}^{[Y]} = G^{[Y]}\left(\rho_{n,k}, \log_{p/q} p^k n\right)\frac{(p^{-\rho_{n,k}}+q^{-\rho_{n,k}})^k n^{-\rho_{n,k}}}{\sqrt{2\pi\beta(\rho_{n,k})k}}\left(1+\mathcal{O}(k^{-1/2})\right),$$

where $G^{[Y]}(\rho,x) = \sum_{j\in\mathbb{Z}}\Gamma(\rho+it_j+1)g^{[Y]}(\rho+it_j)e^{-2\pi ijx}$ is a positive 1-periodic function.

When p decreases and goes to 0.5^+ such that $\alpha_0+\varepsilon \leq k/\log n \leq \alpha_2-\varepsilon$, then the amplitude of oscillations in periodic functions $G^{[X]}(\rho,x)$ and $G^{[Y]}(\rho,x)$ decreases and vanishes. For instance, as we illustrate in Figure 2, the amplitude of oscillations in $G^{[X]}(-0.1,x)$ (right plot) and $G^{[Y]}(-0.1,x)$ (left plot) is a decreasing sequence with respect to $p=0.85$, $p=0.75$, $p=0.65$ and $p=0.55$. For $p=0.65$ and $p=0.55$, we can see that their amplitudes are almost zero, therefore the functions $G^{[X]}(-0.1,x)$ and $G^{[Y]}(-0.1,x)$ are constants. In Table 1, several examples are also given to confirm the following claim.

Theorem 1.2 *Let $\rho_{n,k} := \rho(k/\log n)$ and $\alpha_0 + \varepsilon \le \frac{k}{\log n} \le \alpha_2 - \varepsilon$. Then*

$$\frac{\mu_{n,k}^{[X]}}{\mu_{n,k}^{[Y]}} \overset{n\to\infty}{\sim} \frac{G^{[X]}\left(\rho_{n,k}, \log_{p/q} p^k n\right)}{G^{[Y]}\left(\rho_{n,k}, \log_{p/q} p^k n\right)} \overset{p\to 0.5^+}{\longrightarrow} \frac{\rho_{n,k}(1+\rho_{n,k}) - 8\rho_{n,k}2^{\rho_{n,k}} - 4}{8\rho_{n,k}2^{\rho_{n,k}} - 4\rho_{n,k} - \rho_{n,k}(1+\rho_{n,k})}.$$

Tab. 1: Comparisons of magnitudes for $x \in (0,1)$ and $-2 < \rho(\alpha) < 0$.

Functions of x	p	$\rho(\alpha)$ -0.1	-1	-1.9
$G^{[Y]}(\rho(\alpha), x)$	0.55	0.68	0.44	2.50
	0.75	0.72	0.37	1.81
	0.95	0.73	0.15	0.49
$G^{[X]}(\rho(\alpha), x)$	0.55	8.93	0.56	2.50
	0.75	8.89	0.63	3.10
	0.95	8.85	0.85	4.50
$\frac{G^{[X]}(\rho(\alpha),x)}{G^{[Y]}(\rho(\alpha),x)}$	0.55	13.09	1.27	1.00
	0.75	12.35	1.66	1.62
	0.95	12.12	5.00	9.10

In Theorem 1.3, we derive the asymptotic variances and covariance of the profiles. The methods used to derive these results are the same as the ones used for the expectations. We give examples in Table 2, that the magnitudes of the periodic functions $G_V^{[X]}(\rho(\alpha), x)$, $G_V^{[Y]}(\rho(\alpha), x)$ and $G_C(\rho(\alpha), x)$ increase when $\rho(\alpha)$ grows; and also their amplitudes decrease as $p \to 0.5^+$.

Theorem 1.3 *For some $\varepsilon > 0$, if $\alpha_1 + \varepsilon \le \frac{k}{\log n} \le \alpha_2 - \varepsilon$ and $t_j = 2\pi j/(\log p/q)$ then*

$$\sigma_{n,k}^{[X]^2} = G_V^{[X]}\left(\rho_{n,k}, \log_{p/q} p^k n\right) \frac{(p^{-\rho_{n,k}} + q^{-\rho_{n,k}})^k n^{-\rho_{n,k}}}{\sqrt{2\pi\beta(\rho_{n,k})k}} \left(1 + \mathcal{O}(k^{-1/2})\right),$$
$$\sigma_{n,k}^{[Y]^2} = G_V^{[Y]}\left(\rho_{n,k}, \log_{p/q} p^k n\right) \frac{(p^{-\rho_{n,k}} + q^{-\rho_{n,k}})^k n^{-\rho_{n,k}}}{\sqrt{2\pi\beta(\rho_{n,k})k}} \left(1 + \mathcal{O}(k^{-1/2})\right),$$
$$\gamma_{n,k} = G_C\left(\rho_{n,k}, \log_{p/q} p^k n\right) \frac{(p^{-\rho_{n,k}} + q^{-\rho_{n,k}})^k n^{-\rho_{n,k}}}{\sqrt{2\pi\beta(\rho_{n,k})k}} \left(1 + \mathcal{O}(k^{-1/2})\right),$$

where $\rho_{n,k} = \rho(k/\log n) > -2$ and

$$G_V^{[X]}(\rho, x) = \sum_{j\in\mathbb{Z}} \Gamma(\rho + it_j) g_V^{[X]}(\rho + it_j) e^{-2\pi ijx},$$
$$G_V^{[Y]}(\rho, x) = \sum_{j\in\mathbb{Z}} \Gamma(\rho + it_j + 1) g_V^{[Y]}(\rho + it_j) e^{-2\pi ijx},$$
$$G_C(\rho, x) = \sum_{j\in\mathbb{Z}} \Gamma(\rho + it_j) g_C(\rho + it_j) e^{-2\pi ijx},$$

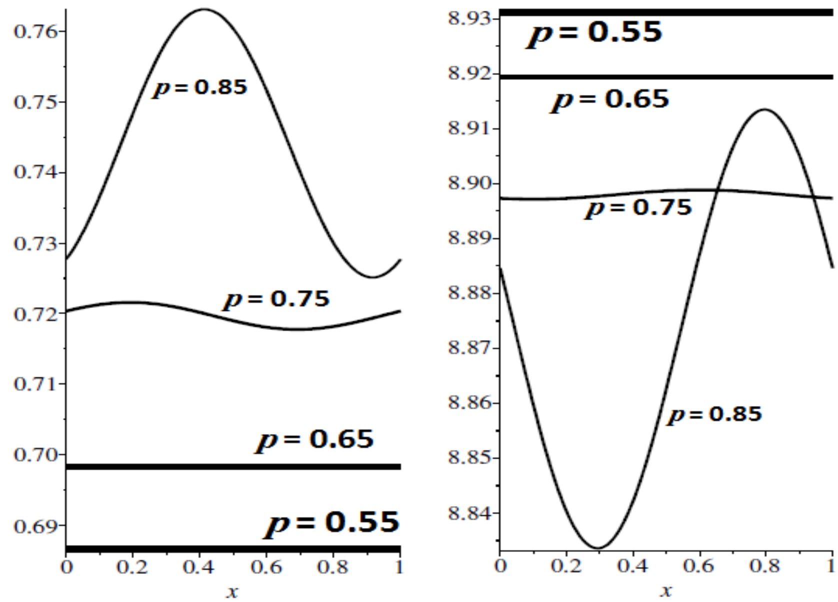

Fig. 2: The fluctuating part of the periodic functions $G^{[Y]}(-0.1, x)$ (left) and $G^{[X]}(-0.1, x)$ (right) for $p = 0.85$, $p = 0.75$, $p = 0.65$, $p = 0.55$ and for $x \in [0, 1]$; their amplitudes tends to zero when $p \to 0.5^+$.

are non-zero 1-periodic functions with

$$\begin{aligned}
g_V^{[X]}(s) &= 1 - 2^{-s} + s\left(p^{-s} + q^{-s} - 2p(p+1)^{-s-1} - 2q(q+1)^{-s-1}\right)\\
&\quad -s(s+1)\left(2^{-s-2}(p^{-s}+q^{-s}) + 3pq - pq2^{-s-1}\right)\\
&\quad +2pqs(s+1)(s+2)\left(p(p+1)^{-s-3} + q(q+1)^{-s-3}\right)\\
&\quad -s(s+1)(s+2)(s+3)p^2q^22^{-s-4},\\
g_V^{[Y]}(s) &= p^{-s} + q^{-s} - 1 + (s+1)\left(2p(p+1)^{-s-2} + 2q(q+1)^{-s-2}\right)\\
&\quad -(s+1)\left(2^{-s-2}(p^{-s}+q^{-s}+1) + 3pq\right) - (s+1)(s+2)(s+3)p^2q^22^{-s-4}\\
&\quad +2pq(s+1)(s+2)\left(p(p+1)^{-s-3} + q(q+1)^{-s-3} - 2^{-s-3}\right),\\
g_C(s) &= s\left(p(p+1)^{-s-1} + q(q+1)^{-s-1}\right) + s(s+1)(s+2)(s+3)p^2q^22^{-s-4}\\
&\quad -s(s+1)\left(p(p+1)^{-s-2} + q(q+1)^{-s-2} - 3pq - 2^{-s-2}(p^{-s}+q^{-s})\right)\\
&\quad -pqs(s+1)(s+2)\left(2p(p+1)^{-s-3} + 2q(q+1)^{-s-3} - 2^{-s-3}\right)\\
&\quad +s\left(1 - 2^{-s-1} - p^{-s} - q^{-s}\right) - s(s+1)pq2^{-s-2}.
\end{aligned}$$

We then prove in Section 3, that both $X_{n,k}$ and $Y_{n,k}$, after proper normalization, are asymptotically bivariate normally distributed for the range $(\alpha_1 + \varepsilon)\log n \le k \le (\alpha_2 - \varepsilon)\log n$.

Theorem 1.4 *For $(\alpha_1+\varepsilon)\log n \le k \le (\alpha_2-\varepsilon)\log n$,*

$$\mathbb{P}\left(\frac{X_{n,k}-\mu_{n,k}^{[X]}}{\sigma_{n,k}^{[X]}} \le x, \frac{Y_{n,k}-\mu_{n,k}^{[Y]}}{\sigma_{n,k}^{[Y]}} \le y\right) = \Phi(x,y;\rho_{n,k}) + o(1), \tag{2}$$

where $\rho_{n,k} := \gamma_{n,k}/\sigma_{n,k}^{[X]}\sigma_{n,k}^{[Y]}$ and $\Phi(x,y;\rho)$ denotes the cumulative distribution function of bivariate standard normal distribution with correlation parameter ρ.

Tab. 2: Comparisons of magnitudes for $x \in (0,1)$ and $\rho(\alpha) > -2$.

Functions of x	p	$\rho(\alpha)$ -1.5	3.5	8.5
$G_V^{[Y]}(\rho(\alpha),x)$	0.55	0.53	225	1.22×10^8
	0.75	0.42	1350	1.58×10^{10}
	0.95	0.13	5×10^5	2.50×10^{16}
$G_V^{[X]}(\rho(\alpha),x)$	0.55	0.51	230	1.23×10^8
	0.75	0.68	1350	1.58×10^{10}
	0.95	0.85	5×10^5	2.50×10^{16}
$G_C(\rho(\alpha),x)$	0.55	-0.05	-220	-1.24×10^8
	0.75	-0.13	-1330	-1.50×10^{10}
	0.95	-0.74	-5×10^5	-3.50×10^{16}

2 Asymptotic Expectations, Variances and Covariance

In this section, except for the proof of positivity of the periodic functions in the asymptotic expansions of the expectations; for the proof of other results, we follow from known methods which were introduced by Park et al. (2009). Namely, we first show that the joint probability generating function of $X_{n,k}$ and $Y_{n,k}$, satisfies a recurrence of the form

$$z_{n,k}(u,w) = \sum_{j=0}^{n}\binom{n}{j}p^j q^{n-j} z_{j,k-1}(u,w) z_{n-j,k-1}(u,w), \qquad (n \ge 0, k \ge 1).$$

From the above recurrence, the expectations, the variances and covariance of $X_{n,k}$ and $Y_{n,k}$ are seen to satisfy a recurrence of the form

$$t_{n,k} = \sum_{j=0}^{n}\binom{n}{j}p^j q^{n-j}(t_{j,k-1}+t_{n-j,k-1}), \qquad (n \ge 0, k \ge 1),$$

with suitable initial conditions. A standard approach is to consider the Poisson transform of $t_{n,k}$, the Poisson generating function $f_k(x) := e^{-x}\sum_n t_{n,k}x^n/n!$, which in turn satisfies the functional equation

$$f_k(x) = f_{k-1}(px) + f_{k-1}(qx). \tag{3}$$

The equation (3) can be solved by a simple iteration argument and has the explicit solution,

$$f_k(x) = \sum_{j=0}^{k} \binom{k}{j} f_0(p^j q^{k-j} x).$$

The asymptotic solution of (3) can be obtained by using the Mellin transform (see Flajolet et al. (1995)). The final step is to invert from the asymptotics of $f_k(x)$ to recover the asymptotics of $t_{n,k}$. This last step is guided by the Poisson heuristic, which roughly states that

if a sequence $\{a_n\}_n$ is "smooth enough", then $a_n \sim e^{-n} \sum_{j\geq 0} a_n n^j / j!$,

where $a_n \sim b_n$ if $\lim_{n\to\infty} a_n/b_n = 1$. This Poisson heuristic is known as analytic depoissonization, when justified by complex analysis and the saddle point method.

2.1 Recurrences for the Poissonized Means and Second Moments

In a random trie of size n, the number of protected nodes $X_{n,k}$ at level $k \geq 1$, can be computed recursively by computing the number for the two subtrees at level $k-1$. For $k = 0$, the root is protected, if and only if neither the left nor the right subtree contains only one string. This leads to the following distributional recurrence for $X_{n,k}$:

$$X_{n,k} \stackrel{d}{=} \begin{cases} X_{B_n,k-1} + X^*_{n-B_n,k-1}, & k \geq 1; \\ 1 - \mathbb{I}_{\{1,n-1\}}(B_n), & k = 0, \end{cases} \qquad (n \geq 2),$$

where $\mathbb{I}_A(\cdot)$ is the indicator function of A, $X_{n,k} \stackrel{d}{=} X^*_{n,k}$, $B_n \stackrel{d}{=} \text{Binomial}(n,p)$ and $X_{n,k}$, $X^*_{n,k}$, B_n are independent. Also, for $k \geq 0$, $X_{0,k} = X_{1,k} = 0$.

Similarly, we have $Y_{0,k} = Y_{1,k} = 0$ for $k \geq 0$, and

$$Y_{n,k} \stackrel{d}{=} \begin{cases} Y_{B_n,k-1} + Y^*_{n-B_n,k-1}, & k \geq 1; \\ \mathbb{I}_{\{1,n-1\}}(B_n), & k = 0, \end{cases} \qquad (n \geq 2),$$

where $Y_{n,k} \stackrel{d}{=} Y^*_{n,k}$ and $Y_{n,k}$, $Y^*_{n,k}$, B_n are independent.

Let $F_{n,k}(u,w) := \mathbb{E}[u^{X_{n,k}} w^{Y_{n,k}}]$ be the joint probability generating function of $X_{n,k}$ and $Y_{n,k}$. Then

$$F_{n,k}(u,w) = \sum_{j=0}^{n} \binom{n}{j} p^j q^{n-j} F_{j,k-1}(u,w) F_{n-j,k-1}(u,w), \qquad (n \geq 0, k \geq 1), \tag{4}$$

with the initial and boundary conditions

$$F_{n,0}(u,w) = \begin{cases} u + n(pq^{n-1} + p^{n-1}q)(w-u), & n \geq 3; \\ u + 2pq(w-u), & n = 2, \\ 1, & n = 0, 1. \end{cases}$$

By taking first and second partial derivatives with respect to u, w on both sides of (4); and then substituting $u = 1$, $w = 1$, we see that $\mu^{[X]}_{n,k}$, $\mu^{[Y]}_{n,k}$, $\nu^{[X]}_{n,k} := \mathbb{E}(X^2_{n,k})$, $\nu^{[Y]}_{n,k} := \mathbb{E}(Y^2_{n,k})$ and $\zeta^{[XY]}_{n,k} := \mathbb{E}(X_{n,k} Y_{n,k})$

satisfy the following recurrences for $n \geq 0, k \geq 1$:

$$\mu_{n,k}^{[X]} = \sum_{j=0}^{n} \binom{n}{j} p^j q^{n-j} \big(\mu_{j,k-1}^{[X]} + \mu_{n-j,k-1}^{[X]}\big), \qquad \mu_{n,k}^{[Y]} = \sum_{j=0}^{n} \binom{n}{j} p^j q^{n-j} \big(\mu_{j,k-1}^{[Y]} + \mu_{n-j,k-1}^{[Y]}\big),$$

$$\nu_{n,k}^{[X]} = \sum_{j=0}^{n} \binom{n}{j} p^j q^{n-j} \big(\nu_{j,k-1}^{[X]} + \nu_{n-j,k-1}^{[X]}\big) + 2\sum_{j=0}^{n} \binom{n}{j} p^j q^{n-j} \mu_{j,k-1}^{[X]} \mu_{n-j,k-1}^{[X]},$$

$$\nu_{n,k}^{[Y]} = \sum_{j=0}^{n} \binom{n}{j} p^j q^{n-j} \big(\nu_{j,k-1}^{[Y]} + \nu_{n-j,k-1}^{[Y]}\big) + 2\sum_{j=0}^{n} \binom{n}{j} p^j q^{n-j} \mu_{j,k-1}^{[Y]} \mu_{n-j,k-1}^{[Y]},$$

$$\zeta_{n,k}^{[XY]} = \sum_{j=0}^{n} \binom{n}{j} p^j q^{n-j} \big(\zeta_{j,k-1}^{[XY]} + \zeta_{n-j,k-1}^{[XY]}\big) + \sum_{j=0}^{n} \binom{n}{j} p^j q^{n-j} \big(\mu_{j,k-1}^{[X]} \mu_{n-j,k-1}^{[Y]} + \mu_{j,k-1}^{[Y]} \mu_{n-j,k-1}^{[X]}\big),$$

with $\mu_{n,0}^{[X]} = \nu_{n,0}^{[X]}$, $\mu_{n,0}^{[Y]} = \nu_{n,0}^{[Y]}$ and $\zeta_{n,0}^{[XY]} = 0$, for $n \geq 0$; and

$$\mu_{n,0}^{[X]} = \begin{cases} 1 - n(pq^{n-1} + p^{n-1}q), & n \geq 3; \\ 1 - 2pq, & n = 2, \\ 0, & n = 0, 1, \end{cases} \qquad \mu_{n,0}^{[Y]} = \begin{cases} n(pq^{n-1} + p^{n-1}q), & n \geq 3; \\ 2pq, & n = 2, \\ 0, & n = 0, 1. \end{cases}$$

It follows that the transforms

$$M_k^{[X]}(x) := \sum_{n\geq 0} \mu_{n,k}^{[X]} \frac{x^n}{n!} e^{-x}, \qquad M_k^{[Y]}(x) := \sum_{n\geq 0} \mu_{n,k}^{[Y]} \frac{x^n}{n!} e^{-x},$$

$$N_k^{[X]}(x) := \sum_{n\geq 0} \nu_{n,k}^{[X]} \frac{x^n}{n!} e^{-x}, \qquad N_k^{[Y]}(x) := \sum_{n\geq 0} \nu_{n,k}^{[Y]} \frac{x^n}{n!} e^{-x},$$

$$Z_k^{[XY]}(x) := \sum_{n\geq 0} \zeta_{n,k}^{[XY]} \frac{x^n}{n!} e^{-x}, \qquad C_k(x) := Z_k^{[XY]}(x) - M_k^{[X]}(x) M_k^{[Y]}(x),$$

$$V_k^{[X]}(x) := N_k^{[X]}(x) - M_k^{[X]}(x)^2, \qquad V_k^{[Y]}(x) := N_k^{[Y]}(x) - M_k^{[Y]}(x)^2,$$

and the transform $\hat{M}_0^{[X]}(x) := 1 - M_0^{[X]}(x)$ satisfy

$$M_k^{[X]}(x) = \sum_{j=0}^{k} \binom{k}{j} M_0^{[X]}(p^j q^{k-j} x) = 2^k - \sum_{j=0}^{k} \binom{k}{j} \hat{M}_0^{[X]}(p^j q^{k-j} x) := 2^k - \hat{M}_k^{[X]}(x), \tag{5}$$

$$M_k^{[Y]}(x) = \sum_{j=0}^{k} \binom{k}{j} M_0^{[Y]}(p^j q^{k-j} x), \qquad C_k(x) = \sum_{j=0}^{k} \binom{k}{j} C_0(p^j q^{k-j} x), \tag{6}$$

$$V_k^{[X]}(x) = \sum_{j=0}^{k} \binom{k}{j} V_0^{[X]}(p^j q^{k-j} x), \qquad V_k^{[Y]}(x) = \sum_{j=0}^{k} \binom{k}{j} V_0^{[Y]}(p^j q^{k-j} x), \tag{7}$$

for $k \geq 1$ with

$$\begin{aligned}
M_0^{[X]}(x) &= 1 - e^{-x} - pxe^{-px} - qxe^{-qx} + pqx^2e^{-x},\\
M_0^{[Y]}(x) &= pxe^{-px} + qxe^{-qx} - xe^{-x} - pqx^2e^{-x},
\end{aligned}$$

$$\begin{aligned}
V_0^{[X]}(x) &= N_0^{[X]}(x) - M_0^{[X]}(x)^2 = M_0^{[X]}(x) - M_0^{[X]}(x)^2\\
&= e^{-x} - e^{-2x} + pxe^{-px} + qxe^{-qx} - 2pxe^{-x(1+p)} - 2qxe^{-x(1+q)} + 2pqx^2e^{-2x}\\
&\quad -3pqx^2e^{-x} - p^2x^2e^{-2px} - q^2x^2e^{-2qx} + 2p^2qx^3e^{-x(1+p)} + 2pq^2x^3e^{-x(1+q)} - p^2q^2x^4e^{-2x},\\
V_0^{[Y]}(x) &= N_0^{[Y]}(x) - M_0^{[Y]}(x)^2 = M_0^{[Y]}(x) - M_0^{[Y]}(x)^2\\
&= pxe^{-px} + qxe^{-qx} + 2px^2e^{-x(1+p)} + 2qx^2e^{-x(1+q)} + 2p^2qx^3e^{-x(1+p)} + 2pq^2x^3e^{-x(1+q)}\\
&\quad -p^2x^2e^{-2px} - q^2x^2e^{-2qx} - 2pqx^3e^{-2x} - p^2q^2x^4e^{-2x} - 3pqx^2e^{-x} - xe^{-x} - x^2e^{-2x},\\
C_0(x) &= -M_0^{[X]}(x)M_0^{[Y]}(x)\\
&= xe^{-x} + pxe^{-(1+p)x} + qxe^{-(1+q)x} + p^2x^2e^{-2px} + q^2x^2e^{-2qx} + 3pqx^2e^{-x} + pqx^3e^{-2x}\\
&\quad -xe^{-2x} - pxe^{-px} - qxe^{-qx} - px^2e^{-(1+p)x} - qx^2e^{-(1+q)x} - pqx^2e^{-2x}\\
&\quad -2pq^2x^3e^{-(1+q)x} - 2p^2qx^3e^{-(1+p)x} + p^2q^2x^4e^{-2x}.
\end{aligned}$$

Consider $\bar{I}_{n,k} := 2^k - I_{n,k}$. By Section 6.2 in Park et al. (2009), for $k \geq 1$,

$$V_k^{[I]}(x) = \sum_{j=0}^{k} \binom{k}{j} V_0^{[I]}(x)(p^j q^{k-j} x), \quad \text{and} \quad V_0^{[I]}(x) = (x+1)e^{-x}(1-(x+1)e^{-x}),$$

where

$$V_k^{[I]}(x) := \sum_{n\geq 0} \mathbb{E}(\bar{I}_{n,k}^2)\frac{x^n}{n!}e^{-x} - \Big(\sum_{n\geq 0} \mathbb{E}(\bar{I}_{n,k})\frac{x^n}{n!}e^{-x}\Big)^2.$$

From this definition, for $k \geq 0$, we obtain

$$2C_k(x) = V_k^{[I]}(x) - V_k^{[X]}(x) - V_k^{[Y]}(x). \tag{8}$$

2.2 Depoissonization

In this section, we first derive the asymptotic approximations to $\mu_{n,k}^{[X]}$ and $\mu_{n,k}^{[Y]}$. We are mainly interested in the behaviour of $M_k^{[X]}(x)$, $M_k^{[Y]}(x)$, for $x = n$, since by analytic depoissonization we expect that $\mu_{n,k}^{[X]} \sim M_k^{[X]}(n)$, $\mu_{n,k}^{[Y]} \sim M_k^{[Y]}(n)$.

Let $\hat{M}_k^{*[X]}(s)$ and $M_k^{*[Y]}(s)$ denote the Mellin transforms

$$\hat{M}_k^{*[X]}(s) = \int_0^\infty \hat{M}_k^{[X]}(x)x^{s-1}dx \quad \text{and} \quad M_k^{*[Y]}(s) = \int_0^\infty M_k^{[Y]}(x)x^{s-1}\mathrm{d}x,$$

so that $\hat{M}_k^{*[X]}(s)$ exists for $s \in \mathbb{C}$ with $\Re(s) > 0$; and $M_k^{*[Y]}(s)$ exists for $s \in \mathbb{C}$ with $\Re(s) > -2$. Then (5) and (6) can be rewritten to

$$\hat{M}_k^{*[X]}(s) = (p^{-s} + q^{-s})^k \hat{M}_0^{*[X]}(s), \qquad M_k^{*[Y]}(s) = (p^{-s} + q^{-s})^k M_0^{*[Y]}(s),$$

with

$$\hat{M}_0^{*[X]}(s) = \Gamma(s)\big(1 + s(p^{-s} + q^{-s}) - s(s+1)pq\big),$$
$$M_0^{*[Y]}(s) = \Gamma(s+1)\big(p^{-s} + q^{-s} - 1 - (s+1)pq\big).$$

Hence, by the inverse Mellin transform (Flajolet et al. (1995)) and definition $\rho := \Re(s)$,

$$M_k^{[X]}(x) = 2^k - \frac{1}{2\pi i}\int_{\rho-i\infty}^{\rho+i\infty} \Gamma(s)\hat{g}^{[X]}(s)(p^{-s}+q^{-s})^k x^{-s}\mathrm{d}s, \qquad \rho > 0, \tag{9}$$

$$M_k^{[Y]}(x) = \frac{1}{2\pi i}\int_{\rho-i\infty}^{\rho+i\infty} \Gamma(s+1)g^{[Y]}(s)(p^{-s}+q^{-s})^k x^{-s}\mathrm{d}s, \qquad \rho > -2, \tag{10}$$

For a complex number s we define the function

$$T(s) := p^{-s} + q^{-s}.$$

Here, we evaluate the integrals (9) and (10) via the saddle point method. Thus, it is natural to choose $\rho = \rho_{n,k}$ as the saddle point of the function

$$T(s)^k n^{-s} = e^{k\log T(s) - s\log n},$$

that is the solution of the equation $\frac{\partial}{\partial s}(k\log T(s) - s\log n) = 0$. Equivalently we must find ρ from

$$\frac{k}{\log n} = \frac{p^{-\rho} + q^{-\rho}}{p^{-\rho}\log(1/p) + q^{-\rho}\log(1/q)}, \tag{11}$$

that is, the only real-valued saddle point $\rho = \rho_{n,k} = \rho(\frac{k}{\log n})$ (see (1)).

The integrands in (9) and (10), also has infinitely many complex-valued saddle points of the form $s_j := \rho + 2\pi i j/(\log p/q)$ $(j = \pm1, \pm2, \ldots)$. This is due to the fact

$$T(\rho + it) = p^{-\rho-it}\left(1 + (\frac{q}{p})^{-\rho-it}\right) = p^{-\rho}\cdot e^{-it\log p}\left(1 + (\frac{q}{p})^{-\rho}\cdot e^{it\log\frac{p}{q}}\right).$$

Now by putting $t = 2\pi j/(\log p/q)$, we have

$$\begin{aligned} T\big(\rho + 2\pi i j/(\log p/q)\big) &= p^{-\rho}\cdot e^{-2\pi i j(\log p)/(\log p/q)}\left(1 + (\frac{q}{p})^{-\rho}\cdot e^{2\pi i j}\right) \\ &= e^{-2\pi i j(\log p)/(\log p/q)}T(\rho). \end{aligned} \tag{12}$$

Consequently the behaviour of $T(s)^k x^{-s}$ around $s = s_j$ is almost the same as that of $T(s)^k x^{-s}$ around $s = \rho$. This phenomenon gives a periodic leading factor in the asymptotics of $M_k^{[X]}(n)$, $M_k^{[Y]}(n)$; and also of $\mu_{n,k}^{[X]}$, $\mu_{n,k}^{[Y]}$.

Proof of Theorem 1.1: By evaluating the integrals (9) and (10) via the saddle point method, the proof is quite identical to that of Theorem 2 in Park et al. (2009) (Lemma 7.5 in Drmota (2009)) with the new functions $\hat{g}^{[X]}(s)$, $g^{[X]}(s)$ and $g^{[Y]}(s)$. It remains to show that $G^{[Y]}(\rho, x)$ and $G^{[X]}(\rho, x)$ are positive.

For $x \in [0,1]$ and $\rho > -2$,

$$\begin{aligned}
G^{[Y]}(\rho,x) &= \sum_{j\in\mathbb{Z}} \Gamma(\rho+it_j+1)\big(p^{-\rho-it_j}+q^{-\rho-it_j}-1-(\rho+it_j+1)pq\big)e^{-2\pi ijx}\\
&= q^{-\rho}\sum_{j\in\mathbb{Z}} \Gamma(\rho+it_j+1)e^{-2\pi ij\left(x+\frac{\log q}{\log p/q}\right)} + p^{-\rho}\sum_{j\in\mathbb{Z}} \Gamma(\rho+it_j+1)e^{-2\pi ij\left(x+\frac{\log p}{\log p/q}\right)}\\
&\quad -\sum_{j\in\mathbb{Z}} \Gamma(\rho+it_j+1)e^{-2\pi ijx} - pq\sum_{j\in\mathbb{Z}} \Gamma(\rho+it_j+2)e^{-2\pi ijx}. \qquad (13)
\end{aligned}$$

Since $\Gamma(\rho+it_j)e^{-2\pi ijx/(\log p/q)}$ is the Fourier transform of $f(j) := (\log p/q)e^{-e^{j(\log p/q)+x}}\left(e^{j(\log p/q)+x}\right)^{\rho}$, for $\rho > 0$, then by Poisson summation formula,

$$\sum_{j\in\mathbb{Z}} \Gamma(\rho+it_j)e^{-2\pi ijx/(\log p/q)} = \log(p/q)\sum_{j\in\mathbb{Z}} e^{-e^{j(\log p/q)+x}}\left(e^{j(\log p/q)+x}\right)^{\rho}. \qquad (14)$$

Therefore, from (13) and (14), for $\rho > -1$ and $x \in [0,1]$, we get

$$\begin{aligned}
G^{[Y]}(\rho,x) &= \log(p/q)\sum_{j\in\mathbb{Z}} \left(pe^{-p(p/q)^{j+x}}(p/q)^{(j+x)(\rho+1)} - pqe^{-(p/q)^{j+x}}(p/q)^{(j+x)(\rho+2)}\right)\\
&\quad + \log(p/q)\sum_{j\in\mathbb{Z}} \left(qe^{-q(p/q)^{j+x}}(p/q)^{(j+x)(\rho+1)} - e^{-(p/q)^{j+x}}(p/q)^{(j+x)(\rho+1)}\right)\\
&= \log(p/q)\sum_{j\in\mathbb{Z}} \big(h_{11}(j+x) - h_{12}(j+x) + h_{21}(j+x) - h_{22}(j+x)\big), \qquad (15)
\end{aligned}$$

where

$$\begin{aligned}
h_{11}(x) &:= pe^{-p(p/q)^x}(p/q)^{x(\rho+1)}, \qquad h_{21}(x) := qe^{-q(p/q)^x}(p/q)^{x(\rho+1)},\\
h_{12}(x) &:= pqe^{-(p/q)^x}(p/q)^{x(\rho+2)}, \qquad h_{22}(x) := e^{-(p/q)^x}(p/q)^{x(\rho+1)}.
\end{aligned}$$

Define $h(x) := pe^{-(p/q)^x}(p/q)^{x(\rho+1)}$. By using the inequality $e^{q(p/q)^x} > q(p/q)^x + 1$, it follows that

$$h_{11}(x) > h_{12}(x) + h(x), \qquad \text{for } x \in \mathbb{R}. \qquad (16)$$

Furthermore, since $p(p/q)^x > 0$ it follows that $e^{p(p/q)^x} > 1$ and consequently $qe^{p(p/q)^x} + p > q + p = 1$ which is equivalent to

$$h_{21}(x) + h(x) > h_{22}(x). \qquad (17)$$

Summing up the inequalities (16) and (17), we get

$$h_{11}(x) - h_{12}(x) + h_{21}(x) - h_{22}(x) > 0, \qquad \text{for } x \in \mathbb{R}.$$

Then the positivity of the function $G^{[Y]}(\rho,x)$ follows by (15), for $\rho > -1$. Also, we have

$$\inf_{\substack{-2<\rho\leq -1\\ 0\leq x\leq 1}} G^{[Y]}(\rho,x) \geq 0.1 > 0.$$

Now, we prove the positivity of the function $\hat{G}^{[X]}(\rho, x)$. For $x \in [0,1]$ and $\rho > 0$,

$$\begin{aligned}
\hat{G}^{[X]}(\rho,x) &= \sum_{j\in\mathbb{Z}} \Gamma(\rho+it_j)\big(1+(\rho+it_j)(p^{-\rho-it_j}+q^{-\rho-it_j})-(\rho+it_j)(\rho+it_j+1)pq\big)e^{-2\pi ijx} \\
&= \sum_{j\in\mathbb{Z}} \Gamma(\rho+it_j)e^{-2\pi ijx} - pq\sum_{j\in\mathbb{Z}}\Gamma(\rho+it_j+2)e^{-2\pi ijx} \\
&+q^{-\rho}\sum_{j\in\mathbb{Z}}\Gamma(\rho+it_j+1)e^{-2\pi ij\left(x+\frac{\log q}{\log p/q}\right)} + p^{-\rho}\sum_{j\in\mathbb{Z}}\Gamma(\rho+it_j+1)e^{-2\pi ij\left(x+\frac{\log p}{\log p/q}\right)} \\
&= \log(p/q)\sum_{j\in\mathbb{Z}}\Big(e^{-(p/q)^{j+x}}(p/q)^{(j+x)\rho} - pqe^{-(p/q)^{j+x}}(p/q)^{(j+x)(\rho+2)}\Big) \\
&+\log(p/q)\sum_{j\in\mathbb{Z}}\Big(pe^{-p(p/q)^{j+x}}(p/q)^{(j+x)(\rho+1)} + qe^{-q(p/q)^{j+x}}(p/q)^{(j+x)(\rho+1)}\Big) \\
&\geq G^{[Y]}(\rho,x) > 0.
\end{aligned}$$

Let $G^*(\rho,x) := \sum_{j\in\mathbb{Z}}\Gamma(\rho+it_j)\big(1+(\rho+it_j)(p^{-\rho-it_j}+q^{-\rho-it_j})\big)e^{-2\pi ijx}$. Then

$$\sup_{\substack{-2<\rho<0\\ 0\leq x\leq 1}} G^*(\rho,x) \leq -0.1 < 0,$$

for $-2<\rho<0$ and

$$\begin{aligned}
\hat{G}^{[X]}(\rho,x) &= G^*(\rho,x) - pq\log(p/q)\sum_{j\in\mathbb{Z}} e^{-(p/q)^{j+x}}(p/q)^{(j+x)(\rho+2)} \\
&\leq -0.1 - pq\log(p/q)\sum_{j\in\mathbb{Z}} e^{-(p/q)^{j+x}}(p/q)^{(j+x)(\rho+2)} < 0.
\end{aligned}$$

Hence, $G^{[X]}(\rho,x) = -\hat{G}^{[X]}(\rho,x) > 0$ for $-2<\rho<0$. □

Proof of Theorem 1.2: When $p \to 0.5^+$, then $t_j := 2\pi j/(\log p/q) \to \infty$ and $\Gamma\big(\rho_{n,k}+it_j\big) \to 0$ for $j \neq 0$. Hence, from Theorem 1.1, we have

$$\frac{\mu_{n,k}^{[X]}}{\mu_{n,k}^{[Y]}} \overset{n\to\infty}{\sim} \frac{G^{[X]}\big(\rho_{n,k}, \log_{p/q} p^k n\big)}{G^{[Y]}\big(\rho_{n,k}, \log_{p/q} p^k n\big)} \overset{p\to 0.5^+}{\longrightarrow} \frac{\Gamma(\rho_{n,k})g^{[X]}(\rho_{n,k})}{\Gamma(\rho_{n,k}+1)g^{[Y]}(\rho_{n,k})} = \frac{g^{[X]}(\rho_{n,k})}{\rho_{n,k}g^{[Y]}(\rho_{n,k})}.$$

Substituting $g^{[X]}(\rho_{n,k})$ and $g^{[Y]}(\rho_{n,k})$, we obtain the result. □

Lemma 2.1 *For $\varepsilon > 0$ and ρ, the solution of the equation (11), define*

$$\rho_0 := \begin{cases} \rho, & \text{if } \rho \geq 1 \text{ and } k \geq \alpha_1(1+\varepsilon)\log n; \\ 1, & \text{if } \rho \leq 1. \end{cases}$$

Then for $l = 0, 1, 2, \ldots$, we have

$$\frac{d^l}{dx^l}M_k^{[X]}(x)\bigg|_{x=ne^{i\theta}} = \mathcal{O}\Big(\rho_0^l n^{-l} M_k^{[X]}(n)\Big), \qquad \frac{d^l}{dx^l}M_k^{[Y]}(x)\bigg|_{x=ne^{i\theta}} = \mathcal{O}\Big(\rho_0^l n^{-l} M_k^{[Y]}(n)\Big). \quad (18)$$

Proof: By the same proof of Lemma 4 in Park et al. (2009), the estimates (18) can be proved. □

From (6) and (7); and similar to (9) and (10), for $\rho > -2$, we have

$$V_k^{[X]}(x) = \frac{1}{2\pi i}\int_{\rho-i\infty}^{\rho+i\infty} \Gamma(s+1) g_V^{[X]}(s)(p^{-s}+q^{-s})^k x^{-s}\mathrm{d}s, \tag{19}$$

$$V_k^{[Y]}(x) = \frac{1}{2\pi i}\int_{\rho-i\infty}^{\rho+i\infty} \Gamma(s) g_V^{[Y]}(s)(p^{-s}+q^{-s})^k x^{-s}\mathrm{d}s, \tag{20}$$

$$C_k(x) = \frac{1}{2\pi i}\int_{\rho-i\infty}^{\rho+i\infty} \Gamma(s) g_C(s)(p^{-s}+q^{-s})^k x^{-s}\mathrm{d}s. \tag{21}$$

Lemma 2.2 *Since $M_k^{[Y]}(x)$, $V_k^{[X]}(x)$, $V_k^{[Y]}(x)$ and $C_k(x)$ are functions with the same following estimate,*

$$\begin{cases} \mathcal{O}(x^2), & \text{as } x \to 0; \\ \mathcal{O}(|x|e^{-q\Re(x)}), & \text{as } x \to \infty,\ |\arg(x)| \le \varepsilon, \end{cases}$$

then uniformly for all $k = k(n) \ge 1$ and $x = ne^{i\theta}$, $|\theta| \le \varepsilon$,

$$V_k^{[X]}(x) = \Theta(M_k^{[Y]}(x)), \qquad V_k^{[Y]}(x) = \Theta(M_k^{[Y]}(x)), \qquad C_k(x) = \Theta(M_k^{[Y]}(x)). \tag{22}$$

Therefore

$${\sigma_{n,k}^{[X]}}^2 = \Theta(\mu_{n,k}^{[Y]}), \qquad {\sigma_{n,k}^{[Y]}}^2 = \Theta(\mu_{n,k}^{[Y]}), \qquad \gamma_{n,k} = \Theta(\mu_{n,k}^{[Y]}). \tag{23}$$

Proof: The assertion in (22) follows from Lemma 8 in Park et al. (2009). By analytic depoissonization we expect $\mu_{n,k}^{[Y]} \sim M_k^{[Y]}(n)$, ${\sigma_{n,k}^{[X]}}^2 \sim V_k^{[X]}(n)$, ${\sigma_{n,k}^{[Y]}}^2 \sim V_k^{[Y]}(n)$ and $\gamma_{n,k} \sim C_k(n)$. This proves (23). □

Proof of Theorem 1.3: By evaluating the integrals (19), (20) and (21) via the saddle point method, the proof is similar to that of Theorem 1.1 with the new functions $g_V^{[X]}(s)$, $g_V^{[Y]}(s)$ and $g_C(s)$. The positivity of $G^{[Y]}(\rho, x)$ in the asymptotic expansions of $\mu_{n,k}^{[Y]}$ has been proved in the proof of Theorem 1.1. Thus, by (23), we have ${\sigma_{n,k}^{[X]}}^2 > 0$ and ${\sigma_{n,k}^{[Y]}}^2 > 0$. □

3 Limiting Joint Distribution

In this section, we prove the limiting joint distribution of $X_{n,k}$ and $Y_{n,k}$ is bivariate normal, for our interesting range, *i.e.* $(\alpha_1 + \varepsilon)\log n \le k \le (\alpha_2 - \varepsilon)\log n$.

Our method of the proof of Theorem 1.4 that is the same method which was used by Park et al. (2009) to prove their univariate limit theorem, is roughly as follows. We start from deriving a closed-form expression for the trivariate generating $F_k(x, u, w) := \sum_{n\ge 0} F_{n,k}(u, w)x^n/n!$ by using the recurrence (4). We then will apply the Cauchy integral representation to prove (2), for which we need (for the analytic depoissonization), a crude estimate for $|F_k(ne^{i\theta}, e^{i\varphi}, e^{i\psi})|$ for $|\theta|$ away from zero, as well as a more precise local expansion when $|\theta|$ is very close to zero.

By (4), we have the functional equation

$$F_k(x,u,w) = F_{k-1}(px,u,w)F_{k-1}(qx,u,w), \qquad (k \geq 1),$$

with the initial condition

$$F_0(x,u,w) = e^x + (1-u)(1+x-e^x) + (w-u)(pxe^{qx} + qxe^{px} - x - pqx^2).$$

By iterating this functional equation, we obtain

$$F_k(x,u,w) = \prod_{0\leq j\leq k} F_0(p^j q^{k-j}x,u,w)^{\binom{k}{j}}, \qquad (k \geq 1). \tag{24}$$

In the proof of Theorem 1.3, we need the following upper bound for the depoissonization procedure.

Proposition 3.1 *Uniformly for $k \geq 1$, $r \geq 0$, $|\theta| \leq \pi$, $|u| = 1$ and $|w| = 1$*

$$|F_k(re^{i\theta},u,w)| \leq e^{r-cr\theta^2}, \tag{25}$$

for some constant $c > 0$ independent of k, r and θ.

Proof: In order to prove the above upper bound, we need the following inequality which holds for $r \geq r_0 \approx 2.9183$, $c_1 := 2/(3\pi^2)$:

$$(2+r+pqr^2)\big(e^{c_1 qr\theta^2/2}+1\big) \leq \left(2+r+\frac{r^2}{4}\right)(e^{r/6}+1) \leq e^r. \tag{26}$$

For $r \leq r_0$, we consider the expansion

$$F_0(x,u,w) = 1 + x + \frac{x^2}{2}\big(u - 2pq(u-w)\big) + \sum_{j\geq 3}\frac{x^j}{j!}\big(u - j(pq^{j-1}+p^{j-1}q)(u-w)\big).$$

Define $\delta_2 := 2pq$ and $\delta_j := j(pq^{j-1}+p^{j-1}q)$, for $j \geq 3$. It is easy to see that $\delta_j \leq 0.75$, for $j \geq 2$ and $0.5 \leq p \leq 1$. Thus

$$\begin{aligned}
\big|F_0(re^{i\theta},e^{i\varphi},e^{i\psi})\big| &\leq |1+re^{i\theta}| + \sum_{j\geq 2}\frac{r^j}{j!}\big|(1-\delta_j)e^{i\varphi} + \delta_j e^{i\psi}\big| \\
&\leq |1+re^{i\theta}| + \sum_{j\geq 2}\frac{r^j}{j!} \\
&\leq e^{r-c_2 r\theta^2}, \qquad \text{(By (76) in Park et al. (2009))},
\end{aligned} \tag{27}$$

uniformly for $1 \leq r \leq r_0$, $|\theta| \leq \pi$ and $c_2 := 2/(\pi^2(1+r_0)^2e^{r_0})$.

Now suppose $r \geq r_0 \approx 2.9183$. We can rewrite $F_0(x,u,w)$ as follows:

$$F_0(x,u,w) = ua_1(px)a_1(qx) + 1 - u + x + w\left(xqa_2(px) + xpa_2(qx)\right) + wpqx^2,$$

where $a_1(x) := e^x - x$ and $a_2(x) := e^x - 1 - x$. By (26) and applying Lemma 6 in Park et al. (2009),

$$\begin{aligned}\left|F_0(re^{i\theta}, e^{i\varphi}, e^{i\psi})\right| &\le a_1(pr)a_1(qr)e^{-c_1 r\theta^2/2} + qra_2(pr)e^{-c_1 pr\theta^2} + pra_2(qr)e^{-c_1 qr\theta^2} + 2 + r + pqr^2 \\ &\le (e^r + 2)e^{-c_1 qr\theta^2} + (2 + r + pqr^2)\left(1 - e^{-c_1 qr\theta^2}\right) \\ &\le e^{r - c_1 qr\theta^2/2}, \qquad \text{(By (26))}. \end{aligned} \tag{28}$$

Collecting the two inequalities (27) and (28), we obtain

$$\left|F_0(re^{i\theta}, e^{i\varphi}, e^{i\psi})\right| \le e^{r - cr\theta^2}, \qquad (c := \min\{c_1 q/2, c_2\}),$$

uniformly for $r \ge 0$, $|\theta| \le \pi$. This implies (25) by (24). □

Now, define

$$Q(x, u, w) := \log\left(e^{-x}F_0(x, u, w)\right) = \log\Big(1 - (1-u)\big(1 - a_3(x) - a_4(x)\big) - (1-w)a_4(x)\Big), \tag{29}$$

where $a_3(x) := e^{-x}(1 + x)$ and $a_4(x) := pxe^{-px} + qxe^{-qx} - xe^{-x} - pqx^2e^{-x}$. Let

$$Q_k(x, u, w) := \sum_{j=0}^{k} \binom{k}{j} Q(p^j q^{k-j}x, u, w) = \log\left(e^{-x}F_k(x, u, w)\right).$$

In the following lemma, we prove that $F_0(re^{i\theta}, e^{i\varphi}, e^{i\psi})$ is away from zero for $r \ge 0$ and $|\theta| \le \varepsilon$, implying that $Q_k(x, u, w)$ is well-defined when $|\arg(x)| \le \varepsilon$. This result is needed for the proof of Proposition 3.2.

Lemma 3.1 *The function $Q_k(re^{i\theta}, u, w)$ is well-defined for $r \ge 0$, $|\theta| \le \varepsilon$, $|u| = 1$ and $|w| = 1$.*

Proof: We first show that

$$A := \left|1 - (1 - e^{i\varphi})\big(1 - a_3(r) - a_4(r)\big) - (1 - e^{i\psi})a_4(r)\right| > 0,$$

for $r \ge 0$, $|\theta| \le \varepsilon$, $|u| = 1$ and $|w| = 1$. By direct calculation, we have

$$\begin{aligned}A^2 &= 1 + va_3(r)^2 + za_4(r)^2 + (v + z - t)a_3(r)a_4(r) - va_3(r) - za_4(r) \\ &\ge 1 + va_3(r)^2 + za_4(r)^2 - ta_3(r)a_4(r) - va_3(r) - za_4(r),\end{aligned}$$

where $v := 2(1 - \cos\varphi)$, $t := 2(1 - \cos\psi)$ and $z := 2(1 - \cos(\varphi - \psi))$. Since

$$\begin{aligned}a_4(r) \le \sup_{\substack{r \ge 0 \\ 0.5 \le p \le 1}} a_4(r) &\le \sup_{r \ge 0}\left(pre^{-pr} + qre^{-qr} - re^{-r}\Big|_{p=\frac{1}{2}}\right) \\ &= \sup_{r \ge 0} re^{-r}(e^{r/2} - 1) \approx 0.52069,\end{aligned}$$

we have

$$A^2 \geq \inf_{\substack{r\geq 0\\ 0.5\leq p\leq 1\\ 0\leq v,t,z\leq 2}} A^2 \geq \inf_{\substack{r\geq 0\\ 0.5\leq p\leq 1}} \Big(1+va_3(r)^2+za_4(r)^2-ta_3(r)a_4(r)-va_3(r)-za_4(r)\Big|_{\substack{v=z=2\\ t=0}}\Big)$$
$$= \inf_{\substack{r\geq 0\\ 0.5\leq p\leq 1}} \big(1+2a_3(r)^2+2a_4(r)^2-2a_3(r)-2a_4(r)\big)$$
$$= 1+2a_3(r)^2+2a_4(r)^2-2(a_3(r)+a_4(r))\Big|_{\substack{a_3(r)=1\\ a_4(r)=0.52069}} \approx 0.50085>0.$$

This proves the lemma when $x=r$; the assertion of the lemma follows from analyticity. $\square$

Proposition 3.2 *For* $|\theta|\leq \theta_0 := n^{-2/5}$, $\varphi = o({\sigma^{[Y]}_{n,k}}^{-2/3})$ *and* $\psi = o({\sigma^{[Y]}_{n,k}}^{-2/3})$

$$F_k(ne^{i\theta},e^{i\varphi},e^{i\psi}) = \exp\Bigg(n-\frac{n}{2}\theta^2+M_k^{[X]}(n)i\varphi+M_k^{[Y]}(n)i\psi-nM_k^{[X]\prime}(n)\theta\varphi-nM_k^{[Y]\prime}(n)\theta\psi$$
$$-\frac{1}{2}V_k^{[X]}(n)\varphi^2-\frac{1}{2}V_k^{[Y]}(n)\psi^2-C_k(n)\varphi\psi+\mathcal{O}(E)\Bigg), \tag{30}$$

where

$$E := n|\theta|^3+\rho_0^2{\sigma^{[X]}_{n,k}}^2|\varphi|\theta^2+\rho_0^2{\sigma^{[Y]}_{n,k}}^2\theta^2|\psi|+\rho_0{\sigma^{[X]}_{n,k}}^2|\theta|\varphi^2+\rho_0{\sigma^{[Y]}_{n,k}}^2|\theta|\psi^2$$
$$+\rho_0\gamma_{n,k}|\theta\varphi\psi|+{\sigma^{[Y]}_{n,k}}^2|\psi^3+\psi\varphi^2+\varphi\psi^2|+{\sigma^{[X]}_{n,k}}^2|\varphi|^3.$$

Proof: We start from the expansion of the function $Q(x,u,w)$ defined in (29),

$$Q(x,u,w) = \begin{cases} \big((\tfrac{1}{2}-pq)(1-u)+pq(1-w)\big)x^2+\mathcal{O}\big(|2-u-w||x|^3\big), & \text{as } x\to 0;\\ (1-u)\big(1+\mathcal{O}(|x|e^{-q\Re(x)})\big), & \text{as } x\to\infty,\ |\arg(x)|\leq\varepsilon. \end{cases}$$

By the above expansion, we have

$$Q_k(x,u,w) = \frac{1}{2\pi i}\int_{\rho-i\infty}^{\rho+i\infty} x^{-s}Q^*(s,u,w)(p^{-s}+q^{-s})^k \mathrm{d}s,$$

where $-2<\rho<0$ and $Q^*(s,u,w) := \int_0^\infty x^{s-1}Q(x,u,w)dx$ is defined for $-2<\Re(s)<0$. Note that

$$Q(x,u,w) = (1-u)\big(1-a_3(x)-a_4(x)\big)+(1-w)a_4(x)-\frac{1}{2}\Big((1-u)^2\big(1-a_3(x)-a_4(x)\big)^2$$
$$+2(1-u)(1-w)\big(1-a_3(x)-a_4(x)\big)a_4(x)+(1-w)^2a_4(x)^2\Big)$$
$$+\hat{Q}(x,u,w)(1-u)^3+\tilde{Q}(x,u,w)(1-u)^2(1-w)$$
$$+\breve{Q}(x,u,w)(1-u)(1-w)^2+\bar{Q}(x,u,w)(1-w)^3, \tag{31}$$

where the exact forms of $\hat{Q}$, $\tilde{Q}$, $\breve{Q}$ and $\bar{Q}$ can be obtained by Taylor's reminder formula and are of less important here. We need instead the estimates (the assumptions in Lemma 8 in Park et al. (2009))

$$\mathcal{O}(\hat{Q}(x,u,w)) = \mathcal{O}(\breve{Q}(x,u,w)) = \mathcal{O}(\tilde{Q}(x,u,w)) = \mathcal{O}(\bar{Q}(x,u,w)) = \mathcal{O}\big(|x|^6\big) = \mathcal{O}\big(|x|^2\big),$$

as $x \to 0$ and

$$\begin{aligned}
\tilde{Q}(x,u,w) &= \mathcal{O}(|x|e^{-q\Re(x)}),\\
\breve{Q}(x,u,w) &= \mathcal{O}(|x|^2 e^{-2q\Re(x)}) = \mathcal{O}(|x|e^{-q\Re(x)}),\\
\bar{Q}(x,u,w) &= \mathcal{O}(|x|^3 e^{-3q\Re(x)}) = \mathcal{O}(|x|e^{-q\Re(x)}),\\
\hat{Q}(x,u,w) &= 1 + \mathcal{O}(|x|e^{-q\Re(x)}),
\end{aligned} \tag{32}$$

as $x \to \infty$ in the sector $\{x : |\arg(x)| \le \varepsilon\}$. By (5), (6), (7) and (8), the expansion (31) gives

$$\begin{aligned}
Q_k(x,u,w) &= (1-u)M_k^{[X]}(x) + (1-w)M_k^{[Y]}(x) + \frac{1}{2}(1-u)^2\big(V_k^{[X]}(x) - M_k^{[X]}(x)\big)\\
&+\frac{1}{2}(1-w)^2\big(V_k^{[Y]}(x) - M_k^{[Y]}(x)\big) + (1-u)(1-w)C_k(x)\\
&+(1-u)^3\hat{Q}_k(x,u,w) + (1-u)^2(1-w)\tilde{Q}_k(x,u,w)\\
&+(1-u)(1-w)^2\breve{Q}_k(x,u,w) + (1-w)^3\bar{Q}_k(x,u,w),
\end{aligned}$$

where $\tilde{Q}_k$, $\breve{Q}_k$ and $\bar{Q}_k$ satisfy in (6); and $\hat{Q}_k$ satisfy in (5).

Applying Lemma 8 in Park et al. (2009) and expansions in (32), we have $\tilde{Q}_k(x,u,w) = \Theta(M_k^{[Y]}(x))$, $\breve{Q}_k(x,u,w) = \Theta(M_k^{[Y]}(x))$ and $\bar{Q}_k(x,u,w) = \Theta(M_k^{[Y]}(x))$; and similarly $\hat{Q}_k(x,u,w) = \Theta(M_k^{[X]}(x))$.

These estimates yield, with $x = ne^{i\theta}$,

$$\begin{aligned}
Q_k(x,u,w) &= (1-u)M_k^{[X]}(x) + (1-w)M_k^{[Y]}(x) + \frac{1}{2}(1-u)^2\big(V_k^{[X]}(x) - M_k^{[X]}(x)\big)\\
&+\frac{1}{2}(1-w)^2\big(V_k^{[Y]}(x) - M_k^{[Y]}(x)\big) + (1-u)(1-w)C_k(x)\\
&+\mathcal{O}\Big(\big|1-u\big|^3\big|M_k^{[X]}(ne^{i\theta})\big|\Big) + \mathcal{O}\Big(\big|(2-u-w)^3 - (1-u)^3\big|\big|M_k^{[Y]}(ne^{i\theta})\big|\Big),
\end{aligned}$$

where the $\mathcal{O}$-term holds uniformly for $|\theta| \le \varepsilon$ and $|1-u| = o(1)$ and $|(2-u-w)^3-(1-u)^3| = o(1)$. Since $\mu_{n,k}^{[X]} \to \infty$ and $\mu_{n,k}^{[Y]} \to \infty$, this leads to (30) by expansions of $M_k^{[X]}(ne^{i\theta})$, $M_k^{[Y]}(ne^{i\theta})$, $V_k^{[X]}(ne^{i\theta})$, $V_k^{[Y]}(ne^{i\theta})$ and $C_k(ne^{i\theta})$ at $\theta = 0$, using the estimates in (18). □

Proof of Theorem 1.4: Recall that $\theta_0 := n^{-2/5}$. By Cauchy's integral formula, (25) and (30), we have

$$\begin{aligned}
\mathbb{E}\left(e^{X_{n,k}i\varphi + Y_{n,k}i\psi}\right) &= \frac{n!}{2\pi i}\int_{|x|=n} x^{-n-1}F_k(x, e^{i\varphi}, e^{i\psi})\mathrm{d}x\\
&= \frac{n!n^{-n}}{2\pi}\int_{|\theta|\le\theta_0} x^{-n-1}F_k(ne^{i\theta}, e^{i\varphi}, e^{i\psi})\mathrm{d}\theta + \mathcal{O}\left(n^{-1/10}e^{-cn^{1/5}}\right)\\
&= \frac{n!n^{-n}}{2\pi}e^{n + M_k^{[X]}(n)i\varphi + M_k^{[Y]}(n)i\psi - \frac{1}{2}V_k^{[X]}(n)\varphi^2 - \frac{1}{2}V_k^{[Y]}(n)\psi^2 - C_k(n)\varphi\psi}\\
&\times\int_{-\theta_0}^{\theta_0} e^{-\frac{n}{2}\theta^2 - nM_k^{[X]\prime}(n)\theta\varphi - nM_k^{[Y]\prime}(n)\theta\psi}\big(1 + \mathcal{O}(E)\big)\mathrm{d}\theta + \mathcal{O}\left(n^{-1/10}e^{-cn^{1/5}}\right),
\end{aligned}$$

since $E \to 0$ in the range of integration and when $\varphi = o(\sigma_{n,k}^{[X]^{-4/5}})$ and $\psi = o(\sigma_{n,k}^{[Y]^{-4/5}})$. Applying Stirling's formula, extending the integration limits to $\pm\infty$, making the change of variables $\theta \mapsto \theta n^{-1/2}$ and the refined estimates (by Proposition 1 in Park et al. (2009)) $\gamma_{n,k} \sim C_k(n) - nM_k^{[X]'}(n)^2 M_k^{[Y]'}(n)^2$, $\sigma_{n,k}^{[X]^2} \sim V_k^{[X]}(n) - nM_k^{[X]'}(n)^2$, $\sigma_{n,k}^{[Y]^2} \sim V_k^{[Y]}(n) - nM_k^{[Y]'}(n)^2$, uniformly for $\varphi = o(\sigma_{n,k}^{[X]^{-2/3}})$ and $\psi = o(\sigma_{n,k}^{[Y]^{-2/3}})$, we obtain

$$\begin{aligned}\mathbb{E}\left(e^{X_{n,k}i\varphi + Y_{n,k}i\psi}\right) &= \frac{1}{\sqrt{2\pi}} e^{M_k^{[X]}(n)i\varphi + M_k^{[Y]}(n)i\psi - \frac{\varphi^2}{2}\left(V_k^{[X]}(n) - nM_k^{[X]'}(n)^2\right) - \frac{\psi^2}{2}\left(V_k^{[Y]}(n) - nM_k^{[Y]'}(n)^2\right)} \\ &\times e^{-\varphi\psi\left(C_k(n) - nM_k^{[X]'}(n)M_k^{[Y]'}(n)\right)} \times \int_{-\infty}^{\infty} e^{-\left(\theta + \sqrt{n}M_k^{[X]'}(n)\varphi + \sqrt{n}M_k^{[Y]'}(n)\psi\right)^2/2} \\ &\times \Bigg(1 + \mathcal{O}\Bigg(\frac{1+|\theta|^3}{\sqrt{n}} + \frac{\theta^2\rho_0^2}{n}\left(\sigma_{n,k}^{[X]^2}|\varphi| + \sigma_{n,k}^{[Y]^2}|\psi|\right) + \frac{|\theta|\rho_0}{\sqrt{n}}\Big(\sigma_{n,k}^{[X]^2}\varphi^2 + \sigma_{n,k}^{[Y]^2}\psi^2 \\ &\qquad + \gamma_{n,k}|\varphi\psi|\Big) + \sigma_{n,k}^{[Y]^2}|\psi^3 + \psi\varphi^2 + \varphi\psi^2| + \sigma_{n,k}^{[X]^2}|\varphi|^3\Bigg)\Bigg)\mathrm{d}\theta, \\ &\longrightarrow \exp\left(\mu_{n,k}^{[X]}i\varphi + \mu_{n,k}^{[Y]}i\psi - \frac{\varphi^2}{2}\sigma_{n,k}^{[X]^2} - \frac{\psi^2}{2}\sigma_{n,k}^{[Y]^2} - \varphi\psi\gamma_{n,k}\right) \\ &\times \Big(1 + \mathcal{O}\Big(\sigma_{n,k}^{[Y]^2}|\psi^3 + \psi\varphi^2 + \varphi\psi^2| + \sigma_{n,k}^{[X]^2}|\varphi|^3\Big)\Big),\end{aligned}$$

which implies the result by Lévy's continuity theorem. □

Acknowledgements

The author wants to thank Michael Drmota for his advice and an anonymous referee for his valuable comments leading to several improvements of the manuscript.

References

G.-S. Cheon and L. W. Shapiro. Protected points in ordered trees. *Appl. Math. Lett.*, 21:516–520, 2008.

L. Devroye and S. Janson. Protected nodes and fringe subtrees in some random trees. *Electron. Commun. Probab*, 19:1–10, 2014.

M. Drmota. *Random trees: An interplay between combinatorics and probability*. Springer Wien New York, Vienna, 2009.

R. R. Du and H. Prodinger. Notes on protected nodes in digital search trees. *Appl. Math. Lett.*, 25: 1025–1028, 2012.

P. Flajolet, X. Gourdon, and P. Dumas. Mellin transforms and asymptotics: harmonic sums. *Theoret. Comput. Sci.*, 144:3–58, 1995.

M. Fuchs and C.-K. Lee. A general central limit theorem for shape parameters of m-ary tries and patricia tries. 21:1–26, 2014.

M. Fuchs, H.-K. Hwang, and V. Zacharovas. An analytic approach to the asymptotic variance of trie statistics and related structures. *Theoret. Comput. Sci.*, 527:1–36, 2014.

M. Fuchs, C.-K. Lee, and G.-R. Yu. On 2-protected nodes in random digital trees. *Theor. Comput. Sci.*, 622:111–122, 2016.

J. Gaither and M. D. Ward. The variance of the number of 2-protected nodes in a trie. In *The Tenth Workshop on Analytic Algorithmics and Combinatorics*, pages 43–51, 2013.

J. Gaither, Y. Homma, M. Sellke, and M. D. Ward. On the number of 2-protected nodes in tries and suffix trees. In *Discrete Math. Theor. Comput. Sci. Proc.*, pages 381–398, 2012.

H.-K. Hwang, M. Fuchs, and V. Zacharovas. Asymptotic variance of random symmetric digital search trees. *Discrete Math. Theor. Comput. Sci.*, 12:103–166, 2012.

D. Knuth. *The art of computer programming. Sorting and searching*. Addison-Wesley, Reading, MA, New York, 1998.

H. M. Mahmoud. *Evolution of random search trees*. John Wiley & Sons Inc., New York, 1992.

P. Nicodème. Average profiles, from tries to suffix-trees. In *Discrete Math. Theor. Comput. Sci. Proc.*, pages 257–266, 2005.

G. Park, H.-K. Hwang, P. Nicodème, and W. Szpankowski. Profiles of tries. *SIAM J. Computing*, 38: 1821–1880, 2009.