Asymmetric Rényi Problem*

July 8, 2015

Abram Magner[†] and Michael Drmota[‡] and Wojciech Szpankowski[§]

Abstract

In 1960 Rényi in his Michigan State University lectures asked for the number of random queries necessary to recover a hidden bijective labeling of n distinct objects. In each query one selects a random subset of labels and asks, which objects have these labels? We consider here an asymmetric version of the problem in which in every query an object is chosen with probability $p > 1/2$ and we ignore "inconclusive" queries. We study the number of queries needed to recover the labeling in its entirety (the *height*), to recover a single element (the *fillup level*), and to recover a randomly chosen element (the *typical depth*). This problem exhibits several remarkable behaviors: the depth D_n converges in probability but not almost surely and while it satisfies the central limit theorem its local limit theorem doesn't hold; the height H_n and the fillup level F_n exhibit phase transitions with respect to p in the second term. For example, in our strongest result we prove that with high probability (whp) we need

$$\log_{1/p} n + \log_{p/(1-p)} \log n + o(\log \log n)$$

queries to recover the entire bijection. This should be compared to its symmetric ($p = 1/2$) counterpart established by Pittel and Rubin, who proved that in this case one requires

$$\log_2 n + \sqrt{2 \log_2 n} + o(\sqrt{\log n})$$

queries. Notice the surprising phase transition at the second term. To obtain these results, we take a unified approach via the analysis of the *profile* defined at level k as the number of elements recovered by the kth query. We first establish new precise asymptotic results for the average and variance, and a central limit law, for the profile in the regime where it grows polynomially with n. These results on their own are quite challenging due to extra dependency introduced by the provision of ignoring inconclusive queries, and they provide distributional information about the depth. We then extend the profile results to the boundaries of the central region, leading to the solution of our problem for the height and fillup level. As a bonus, our analysis implies novel results for random PATRICIA tries, as it turns out that the problem is probabilistically equivalent to the analysis of the height, fillup level, typical depth, and profile of a PATRICIA trie built from n independent binary sequences generated by a biased(p) memoryless source. These results are obtained through tools of analytic combinatorics.

Index Terms: Rényi problem, PATRICIA trie, profile, height, fillup level, analytic combinatorics, Mellin transform, depoissonization.

---

*This work was supported by NSF Center for Science of Information (CSoI) Grant CCF-0939370, the Humboldt Research Foundation, and in addition by NSF Grants CCF-1524312, NSA Grant 130923, and NIH Grant 1U01CA198941-01, and the Austrian Science Foundation FWF Grant No. S9604.

[†]A. Magner is with Department of Computer Science, Purdue University, W. Lafayette, IN 47907, USA (email: anmagner@purdue.edu).

[‡]M. Drmota is with the Institute for Discrete Mathematics and Geometry, TU Wien, A-1040 Wien, Austria, (email: michael.drmota@tuwien.ac.at).

[§]W. Szpankowski is with the Department of Computer Science, Purdue University, IN 47907, USA (e-mail: spa@cs.purdue.edu); also with the Faculty of Electronics, Telecommunications and Informatics, Gdańsk University of Technology, Poland.

1 Introduction

Alfred Rényi was known for posing simple problems whose analyses aren't. In his lectures in the summer of 1960 at Michigan State University, Rényi discussed several problems related to random sets [20]. Among them there was the problem of recovering a labeling of n distinct objects by asking random subset questions of the form "which objects correspond to the labels in the set B?". A sequence of queries corresponds to a refinement of partitions of the set of objects. In this work we consider an asymmetric version of the problem in which in every query each object is selected with probability $p > 1/2$ and we *ignore inconclusive queries* (i.e., those that do not refine every element of the previous partition). We study three parameters of this random process: H_n, the number of such queries needed to recover the entire labeling; F_n, the number needed before any elements are recovered; and D_n, the number needed to recover an element selected uniformly at random. Our objective is to present precise probabilistic estimates of these parameters and to study the distributional behavior of D_n.

The symmetric version (i.e., $p = 1/2$) of the problem was discussed by Pittel and Rubin in [18]. Here is how the authors of [18] formulated it. Given two sets X and A of equal cardinality, the task is to identify the unknown bijective mapping $\phi : X \to A$, where we are allowed to ask questions like, given $B \subseteq A$ what is $\phi^{-1}(B)$? The function ϕ is determined once all elements of the partition of A, hence X, have become singletons. In Rényi's formulation the subsets are selected randomly. In the simplest case, considered in [18], one chooses each element of A to be an element of the query set B independently with probability $p = 1/2$. To make the problem interesting, Pittel and Rubin put the additional constraint that the set B is admissible only if it splits *every* nontrivial element of the current partition. We analyze the problem in the asymmetric case ($p > 1/2$).

Again, the question asked by Rényi brings some surprises. For the symmetric model ($p = 1/2$) Pittel and Rubin [18] were able to prove that the number of necessary queries is with high probability (whp) (see Theorem 1)

$$H_n = \log_2 n + \sqrt{2 \log_2 n} + o(\sqrt{\log n}). \tag{1}$$

In this paper, we re-establish this result using a different approach *and* prove that for $p > 1/2$ the number of queries grows whp as

$$H_n = \log_{1/p} n + \log_{p/q} \log n + o(\log \log n), \tag{2}$$

where $q := 1 - p$. Note an unexpected phase transition in the second term! A similar phase transition occurs in the asymptotics for F_n (see Theorem 1):

$$F_n = \begin{cases} \log_{1/q} n - \log_{1/q} \log \log n + o(\log \log \log n) & p > q \\ \log_2 n - \log_2 \log n + o(\log \log n) & p = q = 1/2. \end{cases} \tag{3}$$

The analysis of D_n reveals more surprises (see Theorem 2). We have $D_n / \log n \to 1/h(p)$ (in probability) where $h(p) := -p \log p - q \log q$, but we do not have almost sure convergence. Moreover, D_n appropriately normalized satisfies a central limit result, but not local limit theorem due to some oscillations discussed below.

We establish these results in a novel way by considering first the *profile* $B_{n,k}$, whose analysis is an open problem of its own. The profile at level k is the number of bijection elements revealed by the kth query. Its study is motivated by the fact that many other parameters, including all of those that we mention here, can be written in terms of it. Indeed, $D_n = \mathbb{E}[B_{n,k}]/n$, $H_n = \max\{k : \ B_{n,k} > 0\}$ and $F_n = \min\{k : \ B_{n,k} > 0\} - 1$.

We now discuss our new results concerning the probabilistic behavior of the profile. We establish precise asymptotic expressions for the expected value and variance of $B_{n,k}$ with $k \sim \alpha \log n$, where, for any fixed $\epsilon > 0$, $\alpha \in (1/\log(1/q) + \epsilon, 1/\log(1/p) - \epsilon)$ (the left and right endpoints of this interval are associated with F_n and H_n, respectively). Specifically, we show that both the mean and the variance are of the same (explicit) polynomial order of growth (with respect to n) (see Theorem 3). More precisely, we show that both expected value and variance grow for $k \sim \alpha \log n$ as

$$H(\rho(\alpha), \log_{p/q}(p^k n)) \; \frac{n^{\beta(\alpha)}}{\sqrt{C \log n}}$$

where $\beta(\alpha) \leq 1$ and $\rho(\alpha)$ are complicated functions of α, C is an explicit constant, and $H(\rho, x)$ is a function that is periodic in x. The oscillations come from infinitely many regularly spaced saddle points that we observe when inverting the Mellin transform of the Poisson generating function of $\mathbb{E}[B_{n,k}]$. Finally, we prove a central limit theorem; that is, $(B_{n,k} - \mathbb{E}[B_{n,k}])/\sqrt{\mathrm{Var}[B_{n,k}]} \to \mathcal{N}(0,1)$ where $\mathcal{N}(0,1)$ represents the standard normal distribution. The expected value analysis of $B_{n,k}$ in the central range gives precise distributional information about D_n, since $\Pr[D_n = k] = \mathbb{E}[B_{n,k}]/n$, and the oscillations in $\mathbb{E}[B_{n,k}]$ are the source of the peculiar behavior of D_n.

In order to use these results to estimate the first and the second order term for H_n and F_n, we need to estimate the mean and the variance of the external profile beyond the range $\alpha \in (1/\log(1/q) + \epsilon, 1/\log(1/p) - \epsilon)$; in particular, for F_n and H_n we need expansions at the left and right side, respectively, of this range. This, it turns out, requires a novel approach and analysis, leading to the announced results on the Rényi problem in (2) and (3).

Having described most of our main results, we mention an important equivalence pointed out by Pittel and Rubin [18]. They observed that their version of the Rényi process resembles the construction of a digital tree known as a PATRICIA trie[1] [11, 22]. In fact, the authors of [18] show that H_n is probabilistically equivalent to the height (longest path) of a PATRICIA trie built from n binary sequences generated independently by a memoryless source with bias p (that is, with a "1" generated with probability p; this is often called the *Bernoulli model with bias p*). It is easy to see that F_n is equivalent to the fillup level (depth of the deepest full level), D_n to the typical depth (depth of a randomly chosen leaf), and $B_{n,k}$ to the external profile of the tree (the number of leaves at level k). Thus, our results on these parameters for the Rényi problem directly lead to novel results on PATRICIA tries. In addition to their use as data structures, PATRICIA tries also arise as combinatorial structures which capture the behavior of various processes of interest in computer science and information theory (e.g., in leader election processes without trivial splits [8] and in the solution to Rényi's problem which we study here [18, 2]).

Now we briefly review known facts about PATRICIA tries and other digital trees when built over n independent strings generated by a memoryless source. Profiles of tries in both the asymmetric and symmetric cases were studied extensively in [15]. The expected profiles of digital search trees in both cases were analyzed in [5], and the variance for the asymmetric case was treated in [9]. Some aspects of trie and PATRICIA trie profiles (in particular, the concentration of their distributions) were studied using probabilistic methods in [4, 3]. The depth in PATRICIA for the symmetric model was analyzed in [2, 11] while for the asymmetric model in [21]. The leading asymptotics for the PATRICIA height for the symmetric Bernoulli model was first analyzed by Pittel [16] (see also [22] for suffix trees). The two-term expression for the height of PATRICIA for the symmetric model was first presented in [18] as discussed above (see also [2]).

The plan for the paper is as follows. In the next section we formulate more precisely our problem and present our main results regarding the external profile, height, fillup level, and depth. Sketches of proofs are provided in the last section.

[1]We recall that a PATRICIA trie is a trie in which non-branching paths are *compressed*; that is, there are no unary paths.

2 Main Results

In this section, we formulate precisely Rényi's problem and present our main results. Our goal is to provide precise asymptotics for three natural parameters of the Rényi problem on n objects with each label in a given query being included with probability $p \geq 1/2$: the number F_n of queries needed to identify a single element of the bijection, the number H_n needed to recover the bijection in its entirety, and the number D_n needed to recover an element of the bijection chosen uniformly at random from the n objects. If one wishes to determine the label for a particular object, these quantities correspond to the best, worst, and average case performance, respectively, of the random subset strategy proposed by Rényi. We call these parameters, the fillup level F_n, the height H_n, and the depth D_n, respectively (these names come from the corresponding quantities in random digital trees). One more parameter is relevant: we can present a unified analysis of our main three parameters F_n, H_n, and D_n via the *profile* $B_{n,k}$, which is the number of elements of the bijection on n items identified by the kth query.

Again, the simple question asked by Rényi brings unexpected surprises. Our analysis reveals several remarkable behaviors: the depth D_n converges in probability but not almost surely and while it satisfies the central limit theorem its local limit theorem doesn't hold. Perhaps most surprisingly, the height H_n and the fillup level F_n exhibit phase transitions with respect to p in the second term. This unusual performance is a consequence of an oscillatory behavior of $B_{n,k}$ for the asymmetric problem (when $p \neq 1/2$) as discussed in Theorem 3.

To begin, we express F_n, H_n, and D_n in terms of $B_{n,k}$:

$$F_n = \min\{k : \ B_{n,k} > 0\} - 1 \qquad H_n = \max\{k : \ B_{n,k} > 0\} \qquad \Pr[D_n = k] = \mathbb{E}[B_{n,k}]/n.$$

Using the first and second moment methods, we can then obtain upper and lower bounds on H_n and F_n in terms of the moments of $B_{n,k}$:

$$\Pr[H_n > k] \leq \sum_{j>k} \mathbb{E}[B_{n,k}], \qquad \Pr[H_n < k] \leq \frac{\mathrm{Var}[B_{n,k}]}{\mathbb{E}[B_{n,k}]^2},$$

and

$$\Pr[F_n > k] \leq \frac{\mathrm{Var}[B_{n,k}]}{\mathbb{E}[B_{n,k}]^2}, \qquad \Pr[F_n < k] \leq \mathbb{E}[B_{n,k}].$$

The analysis of the distribution of D_n reduces simply to that of $\mathbb{E}[B_{n,k}]$.

In the next section, we show that the fillup level F_n and the height H_n have the following precise asymptotic expansions. Both exhibit a surprising phase transition with respect to p in the second term.

Theorem 1 (Asymptotics for F_n and H_n)**.** *With high probability,*

$$H_n = \begin{cases} \log_{1/p} n + \log_{p/q} \log n + o(\log\log n) & p > q \\ \log_2 n + \sqrt{2\log_2 n} + o(\sqrt{\log n}) & p = q \end{cases} \tag{4}$$

and

$$F_n = \begin{cases} \log_{1/q} n - \log_{1/q} \log\log n + o(\log\log\log n) & p > q \\ \log_2 n - \log_2 \log n + o(\log\log n) & p = q \end{cases} \tag{5}$$

for large n.

While the behavior of the fillup level F_n could be anticipated [17] (by comparing it to the corresponding result in a Rényi's problem with inconclusive queries), the performance of the height H_n is totally unexpected and stunning.

Moving to the number of questions D_n needed to identify a random element of the bijection, we have the following theorem that brings more unexpected behaviors (note that due to the evolution process of the random PATRICIA trie, all random variables can be defined on the same probability space).

Theorem 2 (Asymptotics and distributional behavior of D_n)**.** *For $p > 1/2$, the normalized depth $D_n/\log n$ converges in probability to $1/h(p)$ where $h(p) := -p\log p - q\log q$ is the entropy, but not almost surely. In fact,*

$$\liminf_{n\to\infty} D_n/\log n = 1/\log(1/q) \qquad (a.s) \qquad \limsup_{n\to\infty} D_n/\log n = 1/\log(1/p).$$

Furthermore, D_n satisfies a central limit theorem; that is, $(D_n - \mathbb{E}[D_n])/\sqrt{\mathrm{Var}[D_n]} \to \mathcal{N}(0,1)$, where $\mathbb{E}[D_n] \sim \frac{1}{h(p)}\log n$ and $\mathrm{Var}[D_n] \sim c\log n$ where c is an explicit constant. A local *limit theorem does not hold: for $x = O(1)$ and $k = \frac{1}{h}(\log n + x\sqrt{\kappa_*(-1)\log n/h})$, where $\kappa_*(-1)$ is some explicit constant and $h = h(p)$, we obtain*

$$\Pr\left[D_n = k\right] \sim H\left(-1; \log_{p/q} p^k n\right) \frac{e^{-x^2/2}}{\sqrt{2\pi C\log n}}$$

for an oscillating function $H(-1; \log_{p/q} p^k n)$ (see Figure 1) defined in Theorem 3 below and an explicitly known constant C.

Again, the depth exhibits a phase transition: for $p = 1/2$ we have $D_n/\log n \to 1/\log 2$ almost surely, which doesn't hold for $p > 1/2$. We note that some of the results on the depth (namely, the convergence in probability and the central limit theorem) are already known (see [19]), but our contribution is a novel derivation of these facts via the profile analysis.

We now explain our approach to the analysis of the moments of $B_{n,k}$ in appropriate ranges. For this, we take an analytic approach [7, 22]. We first explain it for the analysis relevant to D_n, and then show how to extend it for H_n and F_n. More details can be found in the next section.

We start by deriving a recurrence for the average profile, which we denote by $\mu_{n,k} := \mathbb{E}[B_{n,k}]$. It satisfies

$$\mu_{n,k} = (p^n + q^n)\mu_{n,k} + \sum_{j=1}^{n-1}\binom{n}{j}p^j q^{n-j}(\mu_{j,k-1} + \mu_{n-j,k-1}) \tag{6}$$

for $n \geq 2$ and $k \geq 1$, with some initial/boundary conditions; most importantly, $\mu_{n,k} = 0$ for $k \geq n$ and any n. Moreover, $\mu_{n,k} \leq n$ for all n and k owing to the elimination of inconclusive queries. This recurrence arises from conditioning on the number j of objects that are included in the first query. If $1 \leq j \leq n-1$ objects are included, then the conditional expectation is a sum of contributions from those objects that are included and those that aren't. If, on the other hand, all objects are included or all are excluded from the first potential query (which happens with probability $p^n + q^n$), then the partition element splitting constraint on the queries applies, the potential query is ignored as inconclusive, and the contribution is $\mu_{n,k}$.

The tools that we use to solve this recurrence (for details see [12, 14]) are similar to those of the analyses for digital trees [22] such as tries and digital search trees (though the analytical details differ significantly). We first derive a functional equation for the Poisson transform $\tilde{G}_k(z) = \sum_{m\geq 0}\mu_{m,k}\frac{z^m}{m!}e^{-z}$ of $\mu_{n,k}$, which gives

$$\tilde{G}_k(z) = \tilde{G}_{k-1}(pz) + \tilde{G}_{k-1}(qz) + e^{-pz}(\tilde{G}_k - \tilde{G}_{k-1})(qz) + e^{-qz}(\tilde{G}_k - \tilde{G}_{k-1})(pz).$$

This we write as

$$\tilde{G}_k(z) = \tilde{G}_{k-1}(pz) + \tilde{G}_{k-1}(qz) + \tilde{W}_{k,G}(z), \tag{7}$$

and at this point the goal is to determine asymptotics for $\tilde{G}_k(z)$ as $z \to \infty$ in a cone around the positive real axis. When solving (7), $\tilde{W}_{k,G}(z)$ complicates the analysis because it has no closed-form Mellin transform (see below); we handle it via its Taylor series. Finally, depoissonization [22] will allow us to directly transfer the asymptotic expansion for $\tilde{G}_k(z)$ back to one for $\mu_{n,k}$.

To convert (7) to an algebraic equation, we use the *Mellin transform* [6], which, for a function $f : \mathbb{R} \to \mathbb{R}$ is given by

$$f^*(s) = \int_0^\infty z^{s-1} f(z)\, \mathrm{d}z.$$

Using the Mellin transform identities, we end up with an expression for the Mellin transform $G_k^*(s)$ of $\tilde{G}_k(z)$ of the form

$$G_k^*(s) = \Gamma(s+1)A_k(s)(p^{-s} + q^{-s})^k = \Gamma(s+1)A_k(s)T(s)^k,$$

where $A_k(s)$ is an infinite series arising from the contributions coming from the function $\tilde{W}_{k,G}(z)$. It involves $\mu_{m,j} - \mu_{m,j-1}$ for various m and j (see [12, 13]). Locating and characterizing the singularities of $G_k^*(s)$ then becomes important. We find that, for any k, $A_k(s)$ is entire, with zeros at $s \in \mathbb{Z} \cap [-k, -1]$, so that $G_k^*(s)$ is meromorphic, with possible simple poles at the negative integers less than $-k$. The fundamental strip of $\tilde{G}_k(z)$ then contains $(-k-1, \infty)$. It turns out that the main asymptotic contribution comes from infinite number of saddle points defined by the kernel $T(s) = p^{-s} + q^{-s}$.

We then must asymptotically invert the Mellin transform to recover $\tilde{G}_k(z)$. The Mellin inversion formula for $G_k^*(s)$ is given by

$$\tilde{G}_k(z) = \frac{1}{2\pi i}\int_{\rho-i\infty}^{\rho+i\infty} z^{-s} G_k^*(s)\, \mathrm{d}s = \frac{1}{2\pi i}\int_{\rho-i\infty}^{\rho+i\infty} z^{-s}\Gamma(s+1)A_k(s)T(s)^k\, \mathrm{d}s, \tag{8}$$

where ρ is any real number inside the fundamental strip associated with $\tilde{G}_k(z)$. For k in the range in which the profile grows polynomially (that coincides with the range of interest in our analysis of D_n), we evaluate this integral via the saddle point method [7]. Examining $z^{-s}T(s)^k$ and solving the associated saddle point equation

$$\frac{\mathrm{d}}{\mathrm{d}s}[k \log T(s) - s \log z] = 0,$$

we find an explicit formula (10) for $\rho(\alpha)$, the real-valued saddle point of our integrand. The multivaluedness of the complex logarithm then implies that there are *infinitely many* regularly spaced saddle points on this vertical line, for which we must account (these lead directly to oscillations in the $\Theta(1)$ factor in the final asymptotics for $\mu_{n,k}$). The main challenge in completing the saddle point analysis is then to elucidate the behavior of $\Gamma(s+1)A_k(s)$ for $s \to \infty$ along vertical lines: it turns out that this function inherits the exponential decay of $\Gamma(s+1)$ along vertical lines, and we prove it by splitting the sum defining $A_k(s)$ into two pieces, which decay exponentially for different reasons (the first sum decays as a result of the superexponential decay of $\mu_{m,j}$ for $m = \Theta(j)$, which is outside the main range of interest). We end up with an asymptotic expansion for $\tilde{G}_k(z)$ as $z \to \infty$ in terms of $A_k(s)$.

Finally, we must analyze the convergence properties of $A_k(s)$ as $k \to \infty$. We find that it converges uniformly on compact sets to a function $A(s)$ (which is, because of the uniformity, entire). We then apply Lebesgue's dominated convergence theorem to conclude that we can replace $A_k(s)$ with $A(s)$ in the final asymptotic expansion of $\tilde{G}_k(z)$. All of this yields the following theorem which is proved in [12, 14].

Theorem 3 (Moments and limiting distribution for $B_{n,k}$ for k in the central region)**.** *Let $\epsilon > 0$ be independent of n and k, and fix $\alpha \in \left(\frac{1}{\log(1/q)} + \epsilon, \frac{1}{\log(1/p)} - \epsilon\right)$. Then for $k = k_{\alpha,n} \sim \alpha \log n$:*

(i) *The expected external profile becomes*

$$\mathbb{E}[B_{n,k}] = H(\rho(\alpha), \log_{p/q}(p^k n)) \cdot \frac{n^{\beta(\alpha)}}{\sqrt{2\pi\kappa_*(\rho(\alpha))\alpha \log n}} \left(1 + O(\sqrt{\log n})\right), \tag{9}$$

where

$$\rho(\alpha) = -\frac{1}{\log(p/q)} \log\left(\frac{\alpha\log(1/q) - 1}{1 - \alpha\log(1/p)}\right), \qquad \beta(\alpha) = \alpha\log(T(\rho(\alpha))) - \rho(\alpha), \tag{10}$$

and $\kappa_*(\rho)$ *is an explicitly known function of* ρ*. Furthermore,* $H(\rho, x)$ *(see Figure 1) is a non-zero periodic function with period* 1 *in* x *given by*

$$H(\rho, x) = \sum_{j\in\mathbb{Z}} A(\rho + it_j)\Gamma(\rho + 1 + it_j)e^{-2j\pi i x}, \tag{11}$$

where $t_j = 2\pi j/\log(p/q)$*, and*

$$A(s) = 1 + \sum_{j=1}^{\infty} T(s)^{-j} \sum_{n=j}^{\infty} T(-n)(\mu_{n,j} - \mu_{n,j-1})\frac{\phi_n(s)}{n!}, \tag{12}$$

where $\phi_n(s) = \prod_{j=1}^{n-1}(s+j)$ *for* $n > 1$ *and* $\phi_n(s) = 1$ *for* $n \leq 1$*. Here,* $A(s)$ *is an entire function which is zero at the negative integers.*
(ii) *Variance of the profile is* $\mathrm{Var}[B_{n,k}] = \Theta(\mathbb{E}[B_{n,k}])$.
(iii) *Limiting distribution of the normalized profile is normal, that is,*

$$\frac{B_{n,k} - \mu_{n,k}}{\sqrt{\mathrm{Var}[B_{n,k}]}} \xrightarrow{D} \mathcal{N}(0,1)$$

where $\mathcal{N}(0,1)$ *is the standard normal distribution.*

We should point out that the unusual behavior of D_n in Rényi's problem is a direct consequence of the oscillatory behavior of the profile, which disappears for the symmetric case. Furthermore, for the height and fillup level analyses we need to extend Theorem 3 beyond its original central range for α, as discussed in the next section.

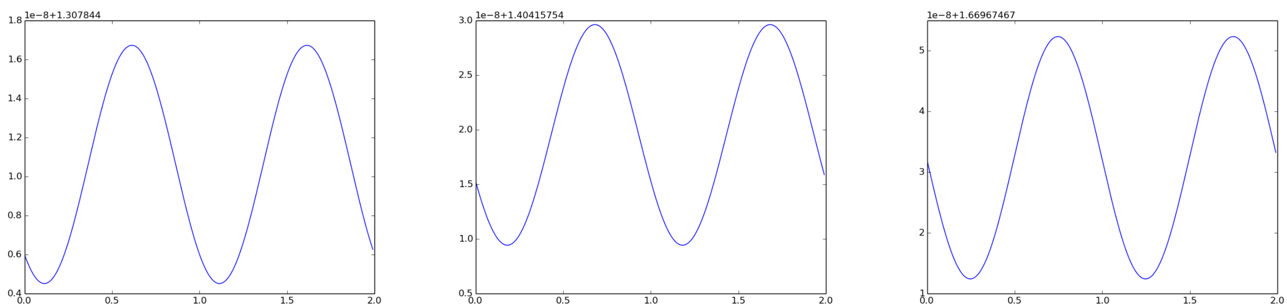

Figure 1: Plots of $H(\rho, x)$ for $\rho = -0.5, 0, 0.5$.

Finally, we note that each of H_n, F_n, D_n, and $B_{n,k}$ corresponds to an analogous parameter of PATRICIA tries built on n independent binary strings generated by a memoryless source with bias $p \geq 1/2$: H_n corresponds to the height (the maximum depth of any leaf), F_n to the fillup level (the highest level k containing 2^k internal nodes), D_n to the typical depth (the depth of a randomly chosen leaf), and $B_{n,k}$ to the external profile (the number of leaves at level k. Thus, the theorems given in this paper also tell us interesting things about random PATRICIA tries.

3 Proof sketches

Now we give sketches of the proofs of Theorems 1 and 2 with more details regarding the proof of Theorem 1 in the Appendix. As stated earlier, the proof of Theorem 3 can be found in [12, 14].

3.1 Sketch of the proof of Theorem 1

To prove our results for H_n and F_n, we extend the analysis of $B_{n,k}$ to the boundaries of the central region (i.e., $k \sim \log_{1/p} n$ and $k \sim \log_{1/q} n$).

Derivation of H_n. Fixing any $\epsilon > 0$, we write, for the lower bound on the height, $k_L = \log_{1/p} n + (1-\epsilon)\psi(n)$ and, for the upper bound, $k_U = \log_{1/p} n + (1+\epsilon)\psi(n)$, for a function $\psi(n) = o(\log n)$ which we are to determine. In order for the first and second moment methods to work, we require

$$\mu_{n,k_L} \xrightarrow{n\to\infty} \infty$$

and

$$\mu_{n,k_U} \xrightarrow{n\to\infty} 0.$$

(We additionally need that $\mathrm{Var}[B_{n,k_L}] = o(\mu^2_{n,k_L})$, but this is not too hard to show by induction using the recurrence for $\tilde{V}_k(z)$, the Poisson variance of $B_{n,k}$.) In order to identify the $\psi(n)$ at which this transition occurs, we define $k = \log_{1/p} n + \psi(n)$, and the plan is to estimate $\mathbb{E}[B_{n,k}]$ via the integral representation (8) for its Poisson transform. Specifically, we consider the inverse Mellin integrand for some $s = \rho \in \mathbb{Z}^- + 1/2$ to be set later. This is sufficient for the upper bound, since, by the exponential decay of the Γ function, the entire integral is at most of the same order of growth as the integrand on the real axis. We expand the integrand in (8), that is,

$$J_k(n,s) := \sum_{j=0}^{k} n^{-s} T(s)^{k-j} \sum_{m\geq j} T(-m)(\mu_{m,j} - \mu_{m,j-1}) \frac{\Gamma(m+s)}{\Gamma(m+1)}, \tag{13}$$

and apply a simple extension of Theorem 2.2, part (iii) of [13] to approximate $\mu_{m,j} - \mu_{m,j-1}$ when $j \to \infty$ and $j \gg \log m$:

Lemma 4 (Precise asymptotics for $\mu_{m,j}$, $j \to \infty$ and m near j)**.** *Let $p \geq q$. For $n \to \infty$ with $k \gg \log n$,*

$$\mu_{n,k} \sim (n-k)^{1/2+\frac{\log q}{\log p}} \frac{n!}{(n-k)!} p^{k^2/2+k/2} q^{k/2} \cdot \exp\left(-\frac{\log^2(n-k)}{2\log(1/p)}\right) \Theta(1). \tag{14}$$

Now, we continue with the evaluation of (13). The jth term of (13) is then of order $p^{\nu_j(n,s)}$, where, defining $\Delta_j = j - \psi(n)$, we set

$$\nu_j(n,s) = \Delta_j^2/2 + \Delta_j(s + \log_{1/p}(1+(p/q)^s) + \psi(n) + 1) - \log_{1/p} n \log_{1/p}(1+(p/q)^s) + \psi(n)^2/2 + o(\psi(n)^2).$$

The factor $T(s)^{k-j}$ ensures that the bounded j terms are negligible. By elementary calculus, we can find the j term which maximizes $\nu_j(n,s)$, and then we minimize over all s, which gives

$$\Delta_j = s + \log_{1/p}(1+(p/q)^s) + \psi(n) + 1, \qquad s = -\psi(n) + O(1).$$

The optimal value for $\nu_j(n,s)$ then becomes

$$\nu_j(n,s) = -\log_{1/p} n \log_{1/p}(1+(p/q)^s) + \psi(n)^2/2 + o(\psi(n)^2). \tag{15}$$

Now, to find $\psi(n)$ for which there is a phase transition, we set the exponent in the above expression equal to zero and solve for $\psi(n)$. When $p = 1/2$, the expression $\log_{1/p}(1 + (p/q)^s)$ becomes 1, which gives

$$-\log_2 n + \psi(n)^2/2(1 + o(1)) = 0 \implies \psi(n) \sim \sqrt{2\log_2 n},$$

as expected. On the other hand, when $p > 1/2$, we cannot solve for $\psi(n)$ directly, owing to the fact that $\log_{1/p}(1 + (p/q)^s)$ now depends on s. We instead observe that the asymptotics of the *Lambert W* function [1] play a key role: by our choice of s, $(p/q)^s = o(1)$, so that

$$\log_{1/p}(1 + (p/q)^s) \sim (p/q)^s / \log(1/p) = e^{s\log(p/q)}/\log(1/p).$$

Then we must solve the equation

$$\log_{1/p} n e^{s\log(p/q)}/\log(1/p) = \psi(n)^2/2.$$

Multiplying by $\log(p/q)$ and taking the square root of both sides gives

$$\sqrt{2\log n}\log(p/q)/\log(1/p) = (\log(p/q)/2)\psi(n)e^{-s\log(p/q)/2},$$

and substituting in our choice of s gives

$$\Theta(\sqrt{\log n}) = (1 + o(1))\frac{\log(p/q)}{2}\psi(n)e^{\psi(n)(1+o(1))\log(p/q)/2}.$$

Setting $W = (1 + o(1))\frac{\log(p/q)}{2}\psi(n)$, this becomes $We^W = \Theta(\sqrt{\log n})$, which is precisely in the form of the recurrence satisfied by the Lambert W function. This yields

$$\psi(n) = \log_{p/q}\log n + O(\log\log\log n).$$

Note that replacing $\psi(n)$ in (15) with $(1 - \epsilon)\psi(n)$ yields as a maximum contribution to the sum

$$p^{-\Theta((\log n)^\epsilon)} \to \infty. \tag{16}$$

while replacing it with $(1 + \epsilon)\psi(n)$ gives

$$p^{\Theta((\log\log n)^2)} \to 0.$$

The above analysis gives asymptotic estimates for $\tilde{G}_k(n)$. We then apply analytic depoissonization [22] to conclude that $\mu_{n,k} \sim \tilde{G}_k(n)$, which gives the claimed result.

Derivation of F_n. We now set $k = \log_{1/q} n + \psi(n)$ and

$$k_L = \log_{1/q} n + (1 + \epsilon)\psi(n), \qquad k_U = \log_{1/q} n + (1 - \epsilon)\psi(n). \tag{17}$$

Here, $\psi(n) = o(\log n)$ is to be determined so as to satisfy $\mu_{n,k_L} \to 0$ and $\mu_{n,k_U} \to \infty$. We use a technique similar to that used in the height proof to determine $\psi(n)$, except now the Γ function asymptotics play a role, since we will choose $\rho \in \mathbb{R}$ tending to ∞. Our first task is to upper bound (as tightly as possible), for each j, the magnitude of the jth term of (13). First, we upper bound

$$T(-m)(\mu_{m,j} - \mu_{m,j-1}) \le 2p^m\mu_{m,j} \le 2p^m m, \tag{18}$$

using the boundary conditions on $\mu_{m,j}$. Next, we apply Stirling's formula to get

$$\begin{aligned}
\frac{\Gamma(m+\rho)}{\Gamma(m+1)} &\sim \sqrt{1+\rho/m}\left(\frac{m+\rho}{e}\right)^{m+\rho}\left(\frac{m+1}{e}\right)^{-(m+1)} && (19)\\
&= e^{(m+\rho)\log(m+\rho)-(m+\rho)+m+1-(m+1)\log(m+1)+O(\log\rho)} && (20)\\
&= \exp((m+\rho)\log(m+\rho) - (m+1)\log(m+1) + O(\rho)) && (21)\\
&= \exp(m\log(m(1+\rho/m)) + \rho\log(\rho(1+m/\rho)) - m\log m - \log m + O(\rho)) && (22)\\
&= \exp(m\log(1+\rho/m) + \rho\log(\rho) + \rho\log(1+m/\rho) - \log m + O(\rho)). && (23)
\end{aligned}$$

Multiplying (18) and (23), then optimizing over all $m \geq j$, we find that the maximum term of the m sum occurs at $m = \rho p/q$ and has a value of

$$\exp(\rho \log \rho + O(\rho)). \tag{24}$$

Now, observe that when $\log m \gg \log \rho$, the contribution of the mth term is $p^{m+o(m)} = e^{-\Theta(m)}$. Thus, setting $j' = \rho^{\log \rho}$ (note that $\log j' = (\log \rho)^2 \gg \log \rho$), we split the m sum into two parts:

$$\sum_{m \geq j} 2p^m m \frac{\Gamma(m+\rho)}{\Gamma(m+1)} = \sum_{m=j}^{j'} 2p^m m \frac{\Gamma(m+\rho)}{\Gamma(m+1)} + \sum_{m=j'+1}^{\infty} 2p^m m \frac{\Gamma(m+\rho)}{\Gamma(m+1)}.$$

The terms of the initial part can be upper bounded by (24), while those of the final part are upper bounded by $e^{-\Theta(m)}$ (so that the final part is the tail of a geometric series). This gives an upper bound of

$$j' e^{\rho \log \rho + O(\rho)} + e^{-\Theta(j')} = e^{(\log \rho)^2 + \rho \log \rho + O(\rho)} = e^{\rho \log \rho + O(\rho)},$$

which holds for any j.

Multiplying this by $n^{-\rho} T(\rho)^{k-j} = q^{\rho \Delta_j + (\Delta_j - \log_{1/q} n) \log_{1/q}(1 + (q/p)^\rho)}$ gives

$$q^{\rho \Delta_j + (\Delta_j - \log_{1/q} n) \log_{1/q}(1+(q/p)^\rho) - \rho \log_{1/q} \rho + O(\rho)}, \tag{25}$$

where Δ_j is again $j - \psi(n)$. Maximizing over the j terms, we find that the largest contribution comes from $j = 0$ (i.e., $\Delta_j = -\psi(n)$). Then, just as in the height upper bound, the behavior with respect to ρ depends on whether or not $p = q$, because $\log_{1/q}(1 + (q/p)^\rho) = 1$ when $p = q$ and is dependent on ρ otherwise. Taking this into account and minimizing over ρ gives that the maximum contribution to the j sum is minimized by setting $\rho = 2^{-\psi(n) - \frac{1}{\log 2}}$ when $p = q$ and $\rho \sim \log_{p/q} \log n$ otherwise. Plugging these choices for ρ into the exponent of (25), setting it equal to 0, and solving for $\psi(n)$ gives $\psi(n) = -\log_2 \log n + O(1)$ when $p = q$ and $\psi(n) \sim -\log_{1/q} \log \log n$ when $p > q$. The evaluation of the inverse Mellin integral with $k = k_L$ as defined in (17) and the integration contour given by $\Re(s) = \rho$ proceeds along lines similar to the height proof, and this yields the desired result.

We remark that the lower bound for F_n may also be derived by relating it to the analogous quantity in regular tries: by definition of the fillup level, there are no unary paths above the fillup level in a standard trie. Thus, when converting the corresponding PATRICIA trie, no path compression occurs above this level, which implies that F_n for PATRICIA is lower bounded by that of tries (and the typical value for tries is the same as in our theorem for PATRICIA). We include the lower bound for F_n via the bounding of the inverse Mellin integral because it is similar in flavor to the corresponding proof of the upper bound (for which no short proof seems to exist).

The upper bound for F_n can similarly be handled by an exact evaluation of the inverse Mellin transform.

3.2 Proof of Theorem 2

Using Theorem 3, we can prove Theorem 2.

Convergence in probability: For the typical value of D_n, we show that

$$\Pr[D_n < (1-\epsilon)1/h(p) \log n] \xrightarrow{n \to \infty} 0, \qquad \Pr[D_n > (1+\epsilon)1/h(p) \log n] \xrightarrow{n \to \infty} 0. \tag{26}$$

For the lower bound, we have

$$\Pr[D_n < (1-\epsilon)1/h(p) \log n] = \sum_{k=0}^{\lfloor (1-\epsilon)1/h(p) \log n \rfloor} \Pr[D_n = k] = \sum_{k=0}^{\lfloor (1-\epsilon)1/h(p) \log n \rfloor} \frac{\mu_{n,k}}{n}.$$

We know from Theorem 3 and the analysis of F_n that, in the range of this sum, $\mu_{n,k} = O(n^{1-\epsilon})$. Plugging this in, we get

$$\Pr[D_n < (1-\epsilon)1/h(p)\log n] = \sum_{k=0}^{\lfloor (1-\epsilon)1/h(p)\log n \rfloor} O(n^{-\epsilon}) = O(n^{-\epsilon}\log n) = o(1).$$

The proof for the upper bound is very similar, except that we appeal to the analysis of H_n instead of F_n.

No almost sure convergence: To show that $D_n/\log n$ does not converge almost surely, we show that

$$\liminf_{n\to\infty} D_n/\log n = 1/\log(1/q), \qquad \limsup_{n\to\infty} D_n/\log n = 1/\log(1/p). \tag{27}$$

For this, we first show that, almost surely, $F_n/\log n \xrightarrow{n\to\infty} 1/\log(1/q)$ and $H_n/\log n \xrightarrow{n\to\infty} 1/\log(1/p)$. Knowing this, we consider the following sequences of events: A_n is the event that $D_n = F_n + 1$, and A'_n is the event that $D_n = H_n$. We note that all elements of the sequences are independent, and $\Pr[A_n], \Pr[A'_n] \geq 1/n$. This implies that $\sum_{n=1}^{\infty} \Pr[A_n] = \sum_{n=1}^{\infty} \Pr[A'_n] = \infty$, so that the Borel-Cantelli lemma tells us that both A_n and A'_n occur infinitely often almost surely (moreover, $F_n < D_n \leq H_n$ by definition of the relevant quantities). This proves (27).

To show the claimed almost sure convergence of $F_n/\log n$ and $H_n/\log n$, we cannot apply the Borel-Cantelli lemmas directly, because the relevant sums do not converge. Instead, we apply a trick which was used in [16]. We observe that both (F_n) and (H_n) are non-decreasing sequences. Next, we show that, on some appropriately chosen subsequence, both of these sequences, when divided by $\log n$, converge almost surely to their respective limits. Combining this with the observed monotonicity yields the claimed almost sure convergence, and, hence, the equalities in (27).

We illustrate this idea more precisely for H_n. By our analysis above, we know that

$$\Pr[|H_n/\log n - 1/\log(1/p)| > \epsilon] = O(e^{-\Theta(\log\log n)^2}).$$

Then we fix t, and we define $n_{r,t} = 2^{t^2 2^{2r}}$. On this subsequence, by the probability bound just stated, we can apply the Borel-Cantelli lemma to conclude that $H_{n_{r,t}}/\log(n_{r,t}) \xrightarrow{r\to\infty} 1/\log(1/p)\cdot(t+1)^2/t^2$ almost surely. Moreover, for every n, we can choose r such that $n_{r,t} \leq n \leq n_{r,t+1}$. Then

$$H_n/\log n \leq H_{n_{r,t+1}}/\log n_{r,t},$$

which implies

$$\limsup_{n\to\infty} \frac{H_n}{\log n} \leq \limsup_{r\to\infty} \frac{H_{n_{r,t+1}}}{\log n_{r,t+1}} \frac{\log n_{r,t+1}}{\log n_{r,t}} = \frac{1}{\log(1/p)} \cdot \frac{(t+1)^2}{t^2}.$$

Taking $t \to \infty$, this becomes $1/\log(1/p)$, as desired. The argument for the lim inf is similar, and this establishes the almost sure convergence of H_n. The derivation is entirely similar for F_n.

Asymptotics for probability mass function of D_n: The asymptotic formula for $\Pr[D_n = k]$ with k as in the theorem follows directly from the fact that $\Pr[D_n = k] = \mathbb{E}[B_{n,k}]/n$, plugging in the expression of Theorem 3 for $\mathbb{E}[B_{n,k}]$.

References

[1] Milton Abramowitz and Irene A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, volume 55 of *National Bureau of Standards Applied Mathematics Series.* For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C., 1964.

[2] Luc Devroye. A note on the probabilistic analysis of patricia trees. *Random Struct. Algorithms*, 3(2):203–214, March 1992.

[3] Luc Devroye. Laws of large numbers and tail inequalities for random tries and patricia trees. *Journal of Computational and Applied Mathematics*, 142:27–37, 2002.

[4] Luc Devroye. Universal asymptotics for random tries and patricia trees. *Algorithmica*, 42(1):11–29, 2005.

[5] Michael Drmota and Wojciech Szpankowski. The expected profile of digital search trees. *J. Comb. Theory Ser. A*, 118(7):1939–1965, October 2011.

[6] Philippe Flajolet, Xavier Gourdon, and Philippe Dumas. Mellin transforms and asymptotics: Harmonic sums. *Theoretical Computer Science*, 144:3–58, 1995.

[7] Philippe Flajolet and Robert Sedgewick. *Analytic Combinatorics.* Cambridge University Press, Cambridge, UK, 2009.

[8] Svante Janson and Wojciech Szpankowski. Analysis of an asymmetric leader election algorithm. *Electronic J. Combin*, 4:1–6, 1996.

[9] Ramin Kazemi and Mohammad Vahidi-Asl. The variance of the profile in digital search trees. *Discrete Mathematics and Theoretical Computer Science*, 13(3):21–38, 2011.

[10] Philippe Jacquet, Charles Knessl, and Wojciech Szpankowski. A note on a problem posed by D. E. Knuth on a satisfiability recurrence. *Combinatorics, Probability, and Computing*, 23, 839-841, 2014.

[11] Donald E. Knuth. *The Art of Computer Programming, Volume 3: (2nd ed.) Sorting and Searching.* Addison Wesley Longman Publishing Co., Inc., Redwood City, CA, USA, 1998.

[12] Abram Magner. *Profiles of PATRICIA Tries.* PhD thesis, Purdue University, December 2015.

[13] Abram Magner, Charles Knessl, and Wojciech Szpankowski. Expected external profile of patricia tries. *Proceedings of the Eleventh Workshop on Analytic Algorithmics and Combinatorics*, pages 16–24, 2014.

[14] Abram Magner and Wojciech Szpankowski. Profile of PATRICIA tries. submitted. `http://www.cs.purdue.edu/~spa/papers/patricia2015.pdf`

[15] G. Park, H. Hwang, P. Nicodème, and W. Szpankowski. Profiles of tries. *SIAM Journal on Computing*, 38(5):1821–1880, 2009.

[16] B. Pittel. Asymptotic growth of a class of random trees. *Ann. Probab.*, 18:414–427, 1985.

[17] B. Pittel, Paths in a random digital tree: limiting distributions, *Adv. in Applied Probability*, 18, 139–155, 1986.

[18] Boris Pittel and Herman Rubin. How many random questions are needed to identify n distinct objects? *Journal of Combinatorial Theory, Series A*, 55(2):292–312, 1990.

[19] B. Rais, P. Jacquet, and W. Szpankowski. Limiting distribution for the depth in PATRICIA tries. *SIAM Journal on Discrete Mathematics*, 6(2):197–213, 1993.

[20] A. Rényi, On Random Subsets of a Finite Set, *Mathematica*, 3, 355-362, 1961.

[21] Wojciech Szpankowski. Patricia tries again revisited. *J. ACM*, 37(4):691–711, October 1990.

[22] Wojciech Szpankowski. *Average Case Analysis of Algorithms on Sequences*. John Wiley & Sons, Inc., New York, NY, USA, 2001.

APPENDIX

3.3 More details of the derivation of H_n

We sketch here in more detail the derivation of the bounds for H_n. We note that the fine analysis of F_n is similar.

Using Lemma 4, we can complete the details of estimating the m sum of the jth term of (13) when $j \to \infty$. We start by upper bounding by

$$\sum_{m \geq j} T(-m)\mu_{m,j},$$

where we've used the fact that $\mu_{m,j} - \mu_{m,j-1} \leq \mu_{m,j}$ and, since we will choose $\rho \in \mathbb{Z}^- + 1/2$, $\Gamma(m+\rho)/\Gamma(m+1) \leq 1$. Next, we split the terms into two parts:

$$\sum_{m \geq j} T(-m)\mu_{m,j} = \sum_{m=j}^{j^3} T(-m)\mu_{m,j} + \sum_{m > j^3} T(-m)\mu_{m,j}.$$

The initial sum is estimable using Lemma 4, since $j \gg m^{1/3}$:

$$\sum_{m=j}^{j^3} T(-m)\mu_{m,j} \leq \sum_{m=j}^{j^3} p^{j^2/2+o(j^2)} \leq j^3 p^{j^2/2+o(j^2)} = p^{j^2/2+o(j^2)}.$$

Meanwhile, the final sum can by upper bounded using the exponential smallness of $T(-m)$ and the fact that $\mu_{m,j} \leq m$:

$$\sum_{m > j^3} T(-m)\mu_{m,j} \leq \sum_{m > j^3} e^{-\Theta(m)} m = \sum_{m > j^3} e^{-\Theta(m)} = e^{-\Theta(j^3)}.$$

Adding these together, we get that the m sum for $j \to \infty$ is $p^{j^2/2+o(j^2)}$, and the rest of the derivation of (15) is as in the height analysis.

To bound those terms for which $j < C$, for any constant C, we trivially upper bound the m sum by

$$\sum_{m \geq j} T(-m)m = O(1),$$

and $n^{-\rho}T(\rho)^{k-j}$ is seen to be negligible because of the factor $T(\rho)^{k-j}$.

3.3.1 Upper bound for H_n

For the upper bound, we show that $\mu_{n,k}$ decays to 0 sufficiently quickly, for $k \geq k_U = \log_{1/p} n + (1+\epsilon)\log_{p/q} \log n$. We set $j_* = j_*(n) = k - \log_{1/p} n$ and $\rho = -j_*(n) + O(1)$, and we note that, since $G_k^*(s)$ is analytic at least in the strip $\Re(s) \in (-k-1, \infty)$, there are no contributions from residues. We bound $|\tilde{G}_k(n)|$ as follows: letting $\mathcal{C}$ denote the vertical line $\Re(s) = \rho$, we split it into a central region (near the real axis) $\mathcal{C}^I$ and tails (bounded away from the real axis) $\mathcal{C}^O$:

$$\mathcal{C}^I = \{\rho + it : \ |t| \leq (\log n)^{(\log\log n)^{1-\delta}}\} \qquad \mathcal{C}^O = \{\rho + it : \ |t| > (\log n)^{(\log\log n)^{1-\delta}}\}, \tag{28}$$

where $\delta < 1$ is some fixed positive number. Intuitively, the tail integral is small because of the exponential decay of the Γ function on vertical lines, and the central region contribution is small by the analysis sketched in the height analysis. We start, in both cases, with the triangle inequality:

$$|\tilde{G}_k(n)| \leq \frac{1}{2\pi} \int_{\rho - i\infty}^{\rho + i\infty} \sum_{j=0}^{k} n^{-\rho} |T(s)^{k-j}| \sum_{m \geq j} T(-m) |\mu_{m,j} - \mu_{m,j-1}| \left| \frac{\Gamma(m+s)}{\Gamma(m+1)} \right| \mathrm{d}s. \tag{29}$$

Bounding the central region: In the central region, we can (essentially) control the integrand by bounding above by its value on the real axis. Multiplying by the length of the central region (which we've chosen to be not too large) gives a sufficient upper bound.

More concretely, we start by noting that

$$|T(s)^{k-j}| = |T(s)|^{k-j} \leq |T(\rho)|^{k-j}, \tag{30}$$

where the inequality follows from writing

$$|T(s)| = |p^{-s}||1 + (p/q)^s| = p^{-\rho}|1 + (p/q)^s| \leq p^{-\rho}(1 + |(p/q)^s|) = p^{-\rho}(1 + (p/q)^\rho) = T(\rho). \tag{31}$$

Here, we've applied the triangle inequality.

Furthermore, it can be checked that $|\Gamma(m+s)| \leq |\Gamma(m+\rho)|$ (which follows easily using the integral representation of the Γ function; see [1]). Applying (30) and the Γ function inequality to (29) gives an upper bound on the integrand of

$$\sum_{j=0}^{k} n^{-\rho} T(\rho)^{k-j} \sum_{m \geq j} T(-m)|\mu_{m,j} - \mu_{m,j-1}| \left| \frac{\Gamma(m+\rho)}{\Gamma(m+1)} \right|.$$

From the analysis of the height, we get that the largest term of this sum is at most $p^{\Theta((\log\log n)^2)}$, so bounding all terms uniformly by this gives an upper bound of

$$(k+1)p^{\Theta((\log\log n)^2)} = p^{\Theta((\log\log n)^2) - \Theta(\log\log n)} = p^{\Theta((\log\log n)^2)}.$$

Since this is a uniform upper bound on the integrand in the central region, to bound the integral, we multiply by the length of the contour, which yields

$$|\mathcal{C}^I| p^{\Theta(\log\log n)^2} = p^{\Theta(\log\log n)^2},$$

since we chose $|\mathcal{C}^I|$ to be $e^{o(\log\log n)^2}$.

Bounding the tails: Here we use the following standard bound on the Γ function: for $s = \rho + it$, provided that $|\mathrm{Arg}(s)|$ is less than and bounded away from π and $|s|$ is sufficiently large, we have

$$|\Gamma(s)| \leq C|t|^{\rho - 1/2} e^{-\pi|t|/2}.$$

This is applicable on $\mathcal{C}^O$, and we again use the fact that $|T(s)| \leq T(\rho)$ and $\mu_{m,j} - \mu_{m,j-1} \leq \mu_{m,j} \leq m$ (justified by the boundary conditions on $\mu_{m,j}$), which yields an upper bound for the integrand of

$$\sum_{j=0}^{k} n^{-\rho} T(\rho)^{k-j} \sum_{m \geq j} T(-m) m \frac{\Theta(|t|^{m+\rho-1/2} e^{-\pi|t|/2})}{\Gamma(m+1)}. \tag{32}$$

Then the m sum becomes

$$C|t|^{\rho-1/2} e^{-\pi|t|/2} \sum_{m \geq j} \frac{m(p|t|)^m}{m!} \leq Cp|t|^{\rho+1/2} e^{-\pi|t|/2} e^{p|t|}, \tag{33}$$

where we've pulled out a factor of $p|t|$, extended the bottom index of the sum to 0, and applied the Taylor series of the exponential function. Note that $-pi/2 + p < 0$, so that we're left with

$$e^{-\Theta(|t|) + (\rho+1/2)\log|t|}.$$

By our choice of $|t|$, this is simply

$$e^{-\Theta(|t|)},$$

uniformly in j. Integrating this on $\mathcal{C}^O$ gives

$$e^{-\Theta((\log n)^{(\log\log n)^{1-\delta}})}.$$

Plugging this upper bound on the integral of the m sum into (29) gives

$$e^{-\Theta((\log n)^{(\log\log n)^{1-\delta}})}\sum_{j=0}^{k} n^{-\rho}T(\rho)^{k-j},$$

and, since $\rho < 0$, $T(\rho)^{k-j} = o(1)$, so that the j sum is upper bounded by

$$kn^{-\rho} = e^{\Theta(\log\log n \log n)},$$

so that the entire integral on the outer tails is at most

$$e^{-\Theta((\log n)^{(\log\log n)^{1-\delta}})}.$$

Summing the contributions on $\mathcal{C}^O$ and $\mathcal{C}^I$: Thus, we have shown that

$$\mu_{n,k} \sim \tilde{G}_k(n) \leq p^{\Theta(\log\log n)^2} + e^{-\Theta((\log n)^{(\log\log n)^{1-\delta}})} = e^{-\Theta(\log\log n)^2}. \tag{34}$$

Now, our original goal was to bound the sum

$$\sum_{k\geq k_U} \mu_{n,k},$$

and the above upper bound is applicable for all terms, but it is too coarse on most of the range. Thus, we split the sum into two parts:

$$\sum_{k=k_U}^{n} \mu_{n,k} = \sum_{k=k_U}^{\lceil(\log n)^2\rceil} \mu_{n,k} + \sum_{k=\lceil(\log n)^2\rceil+1}^{n} \mu_{n,k}.$$

The initial part can be bounded using (34), and the final part we handle using Lemma 4. The location of the split is dictated by two opposing forces: it must be small enough that uniformly upper bounding the initial part by (34) is sufficient and large enough that we can apply Lemma 4 to the tail sum.

The initial sum is then at most

$$e^{-\Theta(\log\log n)^2}\Theta(\log n)^2 = e^{-\Theta(\log\log n)^2+\Theta(\log\log n)} = e^{-\Theta(\log\log n)^2}.$$

The final sum is at most

$$ne^{-\Theta(\log n)^2} = n^{1-\Theta(\log n)}.$$

Adding these upper bounds together shows that

$$\Pr[H_n > k_U] \xrightarrow{n\to\infty} 0,$$

as desired.

3.3.2 Lower bound for H_n

For the lower bound, we first remark that $\mathrm{Var}[B_{n,k}]$ is easily seen to be $O(\mathbb{E}[B_{n,k}])$ by an inductive argument using the Poisson variance $\tilde{V}_k(z)$ (see [12, 14]). Thus, in order to apply the second moment method, it is sufficient to obtain a lower bound for $\mathbb{E}[B_{n,k}]$ for $k = k_L = \log_{1/p} n + (1-\epsilon)\log_{p/q}\log n$. As with the upper bound, we set $j_* = j_*(n) = k - \log_{1/p} n$ and $\rho = -j_*(n) + O(1)$.

First, we rewrite the inverse Mellin integral representation of $\tilde{G}_k(n)$ (recall that this is the Poisson transform of $\mathbb{E}[B_{n,k}]$, and it satisfies $\tilde{G}_k(n) \sim \mathbb{E}[B_{n,k}]$) as follows:

$$\tilde{G}_k(n) = -\sum_{j=0}^{k}\sum_{m\geq j}\frac{T(-m)\mu_{m,j}}{\Gamma(m+1)}\cdot\frac{1}{2\pi i}\int_{\rho-i\infty}^{\rho+i\infty} n^{-s}T(s)^{k-j-1}\Gamma(m+s)\,\mathrm{d}s(1-T(s)),$$

which follows simply from collecting all terms for which $\mu_{m,j}$ is a factor, then exchanging sums and the integral. Note that, by our choice of ρ, $T(s) = o(1)$, and some terms are positive because $\Gamma(m+s)$ is sometimes negative (indeed, the dominant term is positive).

Next, we use the binomial theorem to write $T(s)^{k-j-1}$ as a sum, which gives

$$\tilde{G}_k(n) \sim -\sum_{j=0}^{k}\sum_{m\geq j}\frac{T(-m)\mu_{m,j}}{\Gamma(m+1)}\sum_{\ell=0}^{k-j-1}\binom{k-j-1}{\ell}\frac{1}{2\pi i}\int_{\rho-i\infty}^{\rho+i\infty} n^{-s}(p^{\ell}q^{k-j-\ell-1})^{-s}\Gamma(m+s)\,\mathrm{d}s. \tag{35}$$

We pull $n^m(p^{\ell}q^{k-j-\ell-1})^m$ out of the integral, then make the substitution $s \mapsto s+m$, yielding

$$\frac{1}{2\pi i}\int_{m+\rho-i\infty}^{m+\rho+i\infty} n^{-s}(p^{\ell}q^{k-j-\ell-1})^{-s}\Gamma(s)\,\mathrm{d}s.$$

We can evaluate this exactly via the residue theorem by shifting the integration contour to $\Re(s) = 1$, which gives that the integral is precisely

$$\sum_{\ell'=(-\lceil m+\rho\rceil+1)\,\vee\,0}^{\infty}\frac{(-np^{\ell}q^{k-j-\ell-1})^{\ell'}}{\ell'!},$$

which is a tail of the Taylor series of $e^{-np^{\ell}q^{k-j-\ell-1}}$ (here, $a \vee b$ denotes the maximum of a and b). Now, it can be shown that only those terms of (35) for which $|j - j_*| = O(1)$ contribute non-negligibly to $\tilde{G}_k(n)$. In fact, the dominant term comes from $j = j_*$ and $\lceil m+\rho\rceil = 0$, which gives a maximum contribution of (16). It may be checked that there is not sufficient cancellation to perturb this result.