AofA'10 DMTCS proc. **AM**, 2010, 401–414

Geometric Bucket Trees: Analysis of Linear Bucket Tree

Philippe Jacquet and Paul Mühlethaler

INRIA Rocquencourt, France

We analyse the average number of buckets in a Linear Bucket tree created by n points uniformly dispatched on an interval of length y. A new bucket is created when a point does not fall in an existing bucket. The bucket is the interval of length 2 centered on the point. We illustrate this concept by an interesting tale of how the moon's surface took on its present form. Thanks to an explicit Laplace transform of the Poissonized sequence, and the use of dePoissonization tools, we obtain the explicit asymptotic expansions of the average number of buckets in most of the asymptotic regimes relative to n and y.

Keywords: the keywords are still missing

1 Introduction

Bucket trees are an important concept in data storage [11]. The analysis of the average performance of such trees has been the motivation of numerous seminal papers in the topic of the algorithm analysis [2], [4], [5], [6]. In this paper we introduce a class of data structure that we temporarily call *Geometric Bucket Trees* (GBT) which brings together aspects of discrete mathematics and continuous geometry. It is inspired from the Rényi parking problem [1, 8].

We consider a collection of unit disks called buckets on a large square area, such that no disk contains the center of another disk. Assume that we throw a random point on the square area. Either the point belongs to the existing disks or we create a new disk centered on this new point (see figure 1).

If we replace the square area by a dimension one interval, the bucket disks are replaced by segments and the data structure can be organized like a tree. We call this tree a *linear bucket tree* (LBT).

We consider an interval $[0, y]$ and a sequence $x_1, x_2, \ldots, x_n$ of n points uniformly and independently distributed in that interval. We insert the points in sequence. To insert point x_i in the LBT we run the following algorithm:

- if the tree is empty, then create a bucket with label x_i,
- otherwise if there is a bucket at the root with label x_j $(j < i)$
  - if $|x - x'| < 1$, then store x in the bucket, otherwise
  - if $x < x'$, then go to the left sub-tree, otherwise go the right sub-tree.

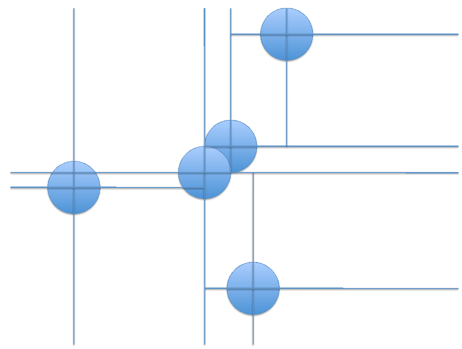

Fig. 1: A geometric bucket tree in dimension 2.

We are not interested in how points are stored inside each bucket (maybe by a list, a trie, a search tree, or even an LBT). We can see that an LBT is a binary tree of buckets. In Figure 2 we display the positioning of seven points. Point x_1 creates the bucket of label x_1, point x_2 creates the bucket of label x_2, x_3 falls in the bucket of label x_1, x_4 and x_5 create two buckets, x_6 falls in the bucket of label x_2 and x_7 creates a bucket.

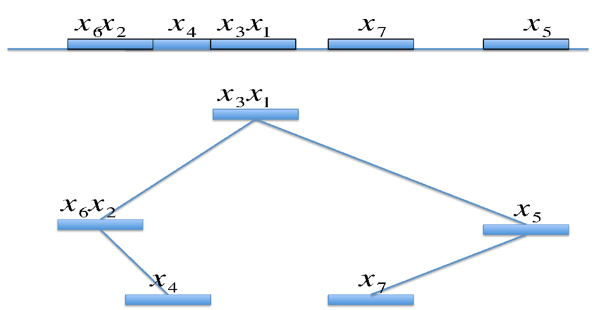

Fig. 2: The linear bucket tree.

We can extend the notion of LBT to any dimension D, by introducing the GBT class which can be similarly defined as a 2^D-ary tree of buckets. In any case when the dimension is two or higher, a GBT would be advantageously complemented as a dag since it is convenient to insert the leaf in several subtrees when the disk overlaps several quarter plans, see for example the leftmost bucket in Figure 1. The shape of the bucket is not important, since it can be a disk or a unit square or a more intricate connected form. For the LBT analysis we propose in this paper, buckets can only be elementary line segments.

The GBT structure is also a way to model wireless transmissions under a *Carrier Sense Multiple Access* (CSMA) strategy. CSMA nodes are located on a square area and transmit after random backoff times that determine their order of transmission. The rule is that a node will actually transmit only if it has not sensed any transmission within its radio range. This strategy is also called ”listen before talk”. More explicitly, when a node actually transmits it draws a disk of exclusion around itself where no other transmission will be allowed. The disk of exclusion is therefore equivalent to the bucket labelled by the coordinate of the transmitter. The average number of possible simultaneously CSMA transmitters is an open problem.

An unexpected application of GBT is in meteoritical science. 3.8 billions years ago the Moon, like the Earth was subject to a "Late Heavy Bombardment" (LHB [9]). At this time huge meteorites punched the planet to the magma and created a crater of lava in an area around twenty times the diameter of the meteorite itself. It is the origin of the lunar bassins (*mare*) on the Moon. Since craters were slowly cooling, if a new meteorite hit this area it would not form a crater as it would be absorbed by the melted lava. Therefore the distribution of the early bassins on the Moon follows a kind of GBT model.

2 Random Linear Bucket Tree, Basic properties, Notations and results

Rényi analyzed the problem of the infinite interval under a Poisson stream of points, therefore we extend here his result to the case where the segment is finite and the number of points is fixed. Let y be a real number and n be an integer. We denote $f_n(y)$ the average number of buckets in an LBT built over n points randomly dispatched over an interval of length y. By convention, when $y \leq 0$ we assume $f_n(y) = 0$. When $0 < y \leq 1$ and $n > 0$, we have $f_n(y) = 1$. By construction the labels of the buckets in an LBT are spaced by 1 or more in distance, therefore $f_n(y) \leq 1 + y$.

We are interested in determining the limit of $f_n(y)$ when $y \to \infty$ and $n \to \infty$. It turns out that the poissonization of function $f_n(y)$ has an explicit expression, and the asymptotic behavior can be characterized. We show among other results that $\lim \frac{1}{y} f_n(y) = \delta$ when $\frac{n}{y} \to \infty$, for $\delta = \int_0^\infty \exp(-2k(\theta))d\theta$, with $k(\theta) = \int_0^\theta \frac{1-e^{-\omega}}{\omega} d\omega$. We have $\delta = 0.7475979203\ldots$, known as Rényi jamming constant.

More precisely when $\liminf \frac{n}{y} > 0$ we show the following expansion:

$$f_n(y) = \delta y + (2\delta - 1) - \frac{e^{-2\gamma}}{n+1}(2y+1) - \frac{e^{-2\gamma}}{(n+2)(n+1)} + O(e^{-\rho y}) + O(e^{-n}) \ .$$

In the case where $\frac{n}{y} \to 0$ we show that

$$f_n(y) = n - \frac{n^2}{y} + \frac{1}{2}\frac{n^2}{y^2} + O(\frac{n^3}{y^2})$$

and the expansion can be continued up to any order.

Unfortunately, when $D > 1$, the asymptotic behavior, even the basic equations, in particular the case $D = 2$, are unknown and are open problems.

The paper is organized as follows. Section 3 describes the basic equation and in particular the use of Poissonization. We compute the first values of $f_n(y)$ for small values of y and n. Section 4 provides some useful lemmas and establish the asymptotic behavior of $f_n(y)$ when y is fixed. This case is not interesting, since it does not give much information beyond the order of magnitudes. Section 5 provides the most interesting results when the parameter y varies and tends to infinity. A surprising result is the fact that the Laplace transform $\tilde{f}(\omega, z)$ of function $f(y, z)$ with respect to variable y has an explicit expression, namely $\frac{e^{2k(\omega)}}{\omega^2} \int_\omega^{\omega+z} e^{-2k(\theta)} d\theta$.

3 Basic equation and poissonization

3.1 Basic equations of sequences

We have the following basic (but nevertheless complicated) equation

$$\begin{aligned} f_{n+1}(y) &= 1+\tfrac{1}{y}\int_0^y \sum_{k+\ell\le n} \tfrac{n!}{k!\ell!(n-k-\ell)!}(\tfrac{x-1}{y})^k(\tfrac{\min\{x-1,1\}+\min\{y-x-1,1\}}{y})^{\ell}(\tfrac{y-1-x}{y})^{n-k-\ell} \\ &\quad \times (f_k(x-1)+f_{n-k-\ell}(y-1-x))\,dx\ . \end{aligned} \tag{1}$$

The explanation of the equation is the following. The first point x_1, splits the segment $[0,y]$ into three parts: $[0,\max\{0,x_1-1\}]$, $[\max\{0,x_1-1\},\min\{y,x_1+1\}]$ and $[\min\{y,x_1+1\},y]$ and we assume that among the n remaining points there are k points on the first segment, ℓ on the second segment and $n-k-\ell$ on the third segment. Since the second segment is the bucket labelled by x_1, those ℓ points can be ignored. The equation can be rewritten:

$$f_{n+1}(y)=1+\frac{2}{y^{n+1}}\int_0^y\sum_k\binom{n}{k}(x-1)^k(y-x+1)^{n-k}f_k(x-1)dx \tag{2}$$

3.2 Sequence Poisson transform

The Poisson transform of sequence $f_n(y)$ is the generating function $\phi(y,z)=\sum_n f_n(y)\frac{z^n}{n!}e^{-z}$. This operation has the consequence to simplify the complicated equation satisfied by the basic sequence in the following functional equation:

$$\frac{\partial}{\partial z}\phi(y,z)+\phi(y,z)=e^{-z}+\frac{2}{y}\int_0^y\phi(x-1,\frac{x-1}{y}z)dx \tag{3}$$

It can be advantageously rewritten with the following normalized generating function $f(y,z)=\phi(y,yz)$ which fixes the parameter $z$bin the right hand side integrand:

$$\frac{1}{y}\frac{\partial}{\partial z}f(y,z)+f(y,z)=1+\frac{2}{y}\int_0^y f(x-1,z)dx\ . \tag{4}$$

Such equation can be solved via Laplace transform: $\tilde{f}(\omega,z)=\int_0^\infty f(y,z)e^{-\omega y}dy$. We will show in section 5 that the Laplace transform $\tilde{f}(\omega,z)=\int_0^{+\infty} f(y,z)e^{-\omega y}$ satisfies the identity, with $k(\omega)=\int_0^\omega\frac{1-e^{-\theta}}{\theta}d\theta$:

$$\tilde{f}(\omega,z)=\frac{e^{2k(\omega)}}{\omega^2}\int_\omega^{\omega+z}e^{2k(\theta)}d\theta\ . \tag{5}$$

We will also use the function $f(y)=\lim_{z\to+\infty}f(y,z)$ which satisfies the functional equation $f(y)=1+\frac{2}{y}\int_0^y f(x-1)dx$ and has Laplace transform $\tilde{f}(\omega)=\frac{e^{2k(\omega)}}{\omega^2}\int_\omega^{+\infty}e^{2k(\theta)}d\theta$, derived from (4) and (5).

3.3 Recursive computing of functions $f(y)$ and $f(y,z)$

We can recursively compute the various values of function $f(y)$ for $y\in[k,k+1]$: Let $f^k(y)=f(k+y)$, we have the recursion:

$$f^{k+1}(y)=1+\frac{k+1}{k+1+y}(f^k(y)-1)+\frac{2}{k+1+y}\int_0^y f^k(x)dx\ . \tag{6}$$

That way we summarize some of these results in the table below:

k	$f^k(y)$
0	1
1	$(3\,y+1)/(y+1)$
2	$(7y+4-4\,\log(y+1))\,/(y+2)$
3	$-1/3\left(-45y-33-48\log(2)+2\,\pi^2+24\text{dilog}(y+2)\right.$ $\left.+60\,\log(y+2)+24\,\log(y+1)\ln(y+2)\right)/(y+3)$

In figure 3 (left) we display the computed function $f(y)$ for y varying in $[0,8]$.

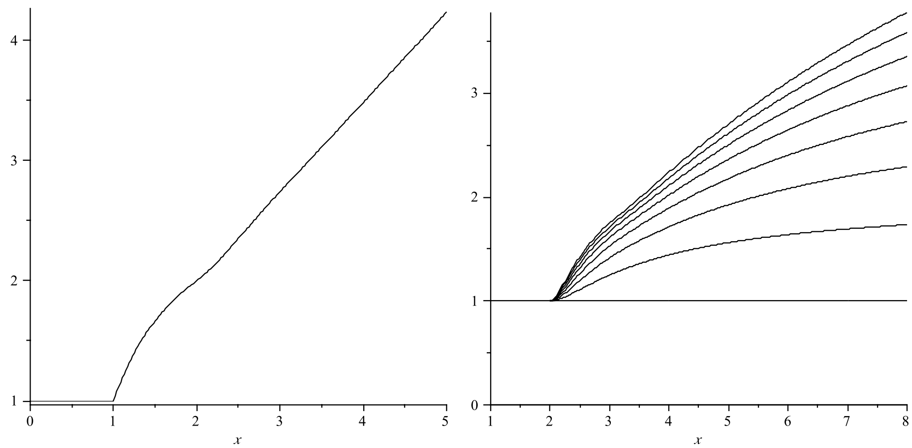

Fig. 3: Left: the computed function $f(y)$. Right: the computed function $f_n(y)$: $f_0(y)=0$ and from bottom to top: $f_1(y)$, $f_2(y)$, $f_3(y)$, $f_4(y)$, $f_5(y)$, $f_6(y)$, $f_7(y)$, $f_8(y)$.

Similarly we can compute the quantity $f^k(y,z)$ with the recursive formula

$$\begin{aligned}\frac{\partial}{\partial z}(f^k(y,z)e^{(y+k)z}) &= (y+k)e^{(y+k)z}+2e^{(y+k)z}\int_0^k f(x-1,z)dx+2e^{(y+k)z}\int_0^y f^{k-1}(x,z)dx\\ &\quad (y+k)e^{(y+k)z}+e^{yz}\frac{\partial}{\partial z}(f^{k-1}(1,z)e^{kz})-ke^{(y+k)z}\\ &\quad +2e^{(y+k)z}\int_0^y f^{k-1}(x,z)dx\end{aligned}$$

we get well $f^0(y,z)=1-e^{-yz}$ and an explicit, but very complicated, formula for $f^1(y,z)$. On figure 3 (right) we display the computed functions $f_n(y)$ for y varying in $[0,8]$. In passing we hint the relatively easy conjecture that for all fixed integer n: $\lim_{y\to\infty} f_n(y)=n$.

4 Basic asymptotic behavior with fixed parameter y

4.1 Asymptotic of the Poisson transform of the sequence $f_n(y)$

Let $F(y,z) = f(y,z)e^{yz}$, from (4) we have the equation

$$\frac{\partial}{\partial z}(F(y,z)) = ye^{yz} + 2e^{yz}\int_0^{y-1} F(x,z)e^{-xz}dx \ . \tag{7}$$

Lemma 1 *Let k be an integer such that $k > 0$, there exists a polynomial $P_k(x)$ of degree $k-1$ with positive coefficients such that the following holds for all z such that for all z such $\Re(z) < 0$ and for all $y \in [0,k]$: $|F(y,z)| \le P_k(|z|)$.*

Proof: We prove the lemma via a recursive argument. The property is true for $k = 1$ since $F(y,z) = e^{yz} - 1$ and in this case we can choose $P_1(x) = 2$. Let us assume that the property is true for an arbitrary integer k and let us take $1 < y < k+1$. We resolve equation (7) like an ordinary differential equation as follows

$$F(y,z) = e^{yz} - 1 + 2z\int_0^1 e^{yzt}dt \int_0^{y-1} F(x,tz)e^{-xzt}dx \ . \tag{8}$$

We have then the inequality between the modulous of the functions

$$|F(y,z)| \le 2 + 2|z| \int_0^1 e^{y\Re(z)t}dt \int_0^{y-1} |F(x,tz)|e^{-x\Re(z)t}dx \tag{9}$$

Since in the left hand-side the variable x is smaller than k we have $|F(x,tz)| \le P_k(|z|t) \le P_k(|z|)$, we get

$$\begin{aligned}
|F(y,z)| &\le 2 + 2|z|P_k(|z|)\int_0^1 e^{y\Re(z)t}dt \int_0^{y-1} e^{-x\Re(z)t}dx \\
&\le 2 + 2|z|P_k(|z|)\frac{e^{y\Re(z)t} - e^{\Re(z)t}}{\Re(z)t} \\
&\le 2 + 2\mu|z|P_k(|z|) \ ,
\end{aligned}$$

with $\mu = \max_{x\ge 0} \frac{1-e^{-x}}{x}$. Therefore the lemma is proven with $P_{k+1}(x) = 2 + 2\mu x P_k(x)$ or, in other words

$$P_k(x) = 2^k \frac{(\mu x)^k - 1}{\mu x - 1} \ . \tag{10}$$

. □

4.2 DePoissonization with fixed parameter y

Let us fix parameter y. For the sequel of the paper $\mathcal{C}$ denote the set of complex numbers z such that $\Re(z) \ge 0$. We show the following theorem

Theorem 1 *For y arbitrary fixed, we have the following estimate:*

$$f_n(y) = f(y, \frac{n}{y}) + O(\frac{1}{n^2}) \tag{11}$$

Proof: We notice that $f_n(y) - f(y)$ is obtained by the dePoissonization of function $f(y, \frac{z}{y}) - f(y)$. We use the "basic dePoissonization" lemma of [7] page 17 (also in [3] and reproduced in [10]) We consider $\mathcal{C}$ as a linear cone with apex 0. We first address condition (O) of dePoissonization: Since $f(y, \frac{z}{y})e^z = F(y, \frac{z}{y})$, from lemma 1 we know that for $\forall z \notin \mathcal{C}$: $|f(y, \frac{z}{y})e^z| \leq P_k(\frac{|z|}{y})$ with $k = \lceil y \rceil$. Thus we got condition (O) for $F(y, \frac{z}{y}) - f(y)e^z$ since for any arbitrary $\alpha > 0$ $P_k(|z|) = o(e^{\alpha|z|})$.

Now we address condition (I). By lemma 3 shown in section 5 we have $\forall z \in \mathcal{C}$: $f(y, z) - f(y) = O(\frac{1}{z})$. Therefore the required condition of the dePoissonization lemma holds with any $\alpha > 0$ and for $\beta = -1$:

$$f_n(y) - f(y) = f(y, \frac{n}{y}) - f(y) + O(\frac{1}{n^2}) \tag{12}$$

□

5 Genaralized asymptotic behavior with varying parameter y

5.1 Generalized Asymptotics via Laplace transform

We introduce a serie of lemmas, that will be useful for the asymptotic analysis. In particular we will focus on the Laplace transform. We introduce the primitive of $f(y, z)$: $g(y, z) = \int_0^y f(x, z)dx$:

$$\frac{\partial^2}{\partial y \partial z} g(y, z) + y \frac{\partial}{\partial y} g(y, z) = y + 2g(y - 1, z) \ . \tag{13}$$

We introduce the Laplace transform $\tilde{g}(\omega, z) = \int_0^\infty g(y, z)e^{-\omega y}dy$. We know that $g_n(y) \leq \frac{y^2}{2} + y$ for all n. Therefore $|g(y, z)| \leq (\frac{y^2}{2} + y)\exp(y|z| - \Re(z)y)$ and the Laplace transform is basically defined for all ω such that $\Re(\omega) > 2|z|$.

Theorem 2 *The Laplace transform $\tilde{g}(\omega, z)$ is defined for all $\Re(\omega) > 0$ and we have the identity*

$$\tilde{g}(\omega, z) = \frac{e^{2k(\omega)}}{\omega^3} \int_\omega^{\omega + z} e^{-2k(\theta)} d\theta \ , \tag{14}$$

with $k(\omega) = \int_0^\omega \frac{1 - e^{-\theta}}{\theta} d\theta = \gamma + \log(\omega) + \mathrm{Ei}(\omega)$, the latter term is the exponential integral function and γ is the Euler-Mascheroni constant.

Proof:

We shall work whith the function $G(y, z) = g(y, z)e^{yz}$, we notice that $G(y, \frac{z}{y}) = \phi(y, z)e^z$, that is the exponential generating function of the sequence $f_n(y)$. The equation (13) becomes

$$\frac{\partial^2}{\partial y \partial z} G(y, z) - z \frac{\partial}{\partial z} G(y, z) - G(y, z) = ye^{yz} + 2e^z G(y - 1, z) \ . \tag{15}$$

Here we introduce the Laplace transform $\tilde{G}(\omega, z)$ of $G(y, \nu)$ with respect to variable y, defined for $\Re(\omega) > |z|$). In passing we have $\tilde{G}(\omega, z) = \tilde{g}(\omega - z, z)$. It satisfies the differential equation:

$$(\omega - z)\frac{\partial}{\partial z}\tilde{G}(\omega, z) - \tilde{G}(\omega, z) = \frac{1}{(\omega - z)^2} + 2e^{z-\omega}\tilde{G}(\omega, z) \ . \tag{16}$$

First, we focus on the Kernel of equation (16) $H(\omega, z)$:

$$\frac{\partial}{\partial z}H(\omega, z) = \frac{1 + 2e^{z-\omega}}{\omega - z}H(\omega, z) \ . \tag{17}$$

which resolves in $H(\omega, z) = \frac{1}{(\omega-z)^3}e^{2k(\omega-z)}$.

We use the fact that $G(y, 0) = 0$, thus $\tilde{G}(\omega, 0) = 0$, the equation (19) has solution:

$$\tilde{G}(\omega, z) = \frac{e^{2k(\omega-z)}}{(\omega - z)^3}\int_0^z e^{-2k(\omega-\theta)}d\theta \tag{18}$$

Second,

$$\frac{\partial}{\partial z}\left(\frac{\tilde{G}(\omega, z)}{H(\omega, z)}\right) = e^{-2k(\omega-z)} \ . \tag{19}$$

Since $\tilde{G}(\omega, z) = \tilde{g}(\omega - z, z)$, we get the claimed result: □

For the following we define the function $f(y) = \lim_{z\to+\infty} f(y, z) = \lim_{z\to+\infty} \phi(y, z)$, which satisfies the equation

$$f(y) = 1 + \frac{2}{y}\int_0^y f(x - 1)dx \ . \tag{20}$$

Lemma 2 *The Laplace transform of $f(y)$ is $\omega\tilde{g}(\omega)$ with*

$$\tilde{g}(\omega) = \frac{e^{2k(\omega)}}{\omega^3}\int_\omega^{+\infty} e^{-2k(\theta)}d\theta \ . \tag{21}$$

and for all $\rho > 0$ the asymptotic evaluation when $y \to \infty$:

$$f(y) = \delta y + 2\delta - 1 + O(e^{-\rho y}) \ , \tag{22}$$

with $\delta = \int_0^\infty e^{-2k(\theta)}d\theta = 0.7475979203\ldots$.

Proof: The equation (21) comes directly from (14). By inverse Laplace transform

$$f(y) = \frac{1}{2i\pi}\int \omega\tilde{g}(\omega)e^{\omega y}d\omega \ , \tag{23}$$

with the integral path being parallel to the imaginary axis in the definition domain of $\tilde{g}(\omega, z)$. In fact it is more convenient to take the Laplace transform of $f(z) - 1$ which is $\frac{e^{2k(\omega)}}{\omega^2}\eta(\omega) - \frac{1}{\omega}$ and is absolutely integrable on any path parallel on the imaginary axis but for the lack of room we omit this rather technical boring proof. There is a double pole on $\omega = 0$. We move the integral path over this pole on $\Re(\omega) - \rho$ for some $\rho > 0$. On this axis we have then

$$f(y) = r(y) + O(e^{-\rho y}) \tag{24}$$

where $r(y)$ is the residue of $\omega\tilde{g}(\omega)e^{\omega y}$ on pole at $\omega = 0$. This residues is equal to $y\delta + 2\delta - 1$ with $\delta = \int_0^\infty e^{-2k(\theta)} d\theta$. □

Lemma 3 *Uniformly for z such that $\Re(z) \geq 0$ complex and for $y > 0$ the following holds:*

$$f(y, z) = f(y) - \eta(z)y - 2\eta(z) - \eta'(z) + O(\frac{e^{-\rho y}}{|z|}) \tag{25}$$

for some $\rho > 0$, with $\eta(z) = \int_z^{+\infty} e^{-2k(\theta)} d\theta$.

Proof: We consider the Laplace transform of $f(y, z) - f(y) + e^{-yz}$ which is equal to $-\frac{e^{2k(\omega)}}{\omega^2}\eta(z+\omega) + \frac{1}{\omega+z}$, and by inverse Laplace transform:

$$f(y) - f(y, z) - e^{-yz} = \frac{1}{2i\pi} \int \left(\frac{e^{2k(\omega)}}{\omega^2}\eta(\omega + z) - \frac{1}{\omega + z} \right) e^{\omega y} d\omega \;, \tag{26}$$

Via a singularity analysis we can move the integral path over the singularity at $\omega = 0$ and fix it on $\Re(\omega) = -\rho$ for some $\rho > 0$. We omit the proof that function $\tilde{g}(\omega, z) - \frac{1}{\omega+z}$ is absolutely integrable on any axis $\Re(\omega) = -\rho \neq 0$ and thus

$$\int_{\Re(\omega)=-\rho} \left(\tilde{f}(\omega, z) - \frac{1}{\omega + z} \right) e^{\omega y} = O(\frac{e^{-\rho y}}{|z|}) \;. \tag{27}$$

. Therefore

$$f(y) - f(y, z) = -r(y, z) + O(\frac{\exp(-\rho y - y\Re(z))}{|z|^3}) + e^{-yz} \tag{28}$$

where $r(y, z)$ is the residues of function $\frac{e^{2k(\omega)}}{\omega^2}e^{\omega y}\eta(z + \omega)$ on $\omega = 0$ (the function $\frac{e^{\omega y}}{\omega+z}$ has no residue on $\omega = 0$). We have the expression:

$$r(y, z) = y\eta(z) + 2\eta(z) + \eta'(z) \;, \tag{29}$$

which terminates the proof. We notice that $\eta'(z) = -e^{-2k(z)}$. □ In passing we get the original following result obtained by Rényi [1] for car parking in an infinite interval. When z is real positive, it is equivalent to the density of a Poisson point process.

Corollary 1 *When $y \to \infty$ and $z > 0$ we have*

$$\lim_{y\to\infty} \frac{f(y, z)}{y} = \int_0^z e^{-2k(\theta)} d\theta \;. \tag{30}$$

Lemma 4 *When $\Re(z) \to +\infty$ we have the following estimate*

$$f(y,z) = f(y) - (y+2)\frac{e^{-2\gamma}}{z} + \frac{e^{-2\gamma}}{z^2} + O(\frac{e^{-\rho y}}{\Re(z)}) + O(e^{-\Re(z)}) \tag{31}$$

Proof: Similar proof as the previous proof, but where we use the fact that $\exp(\Re(z)) \to 0$ and also the fact that $e^{-2k(z)} = \frac{e^{-2\gamma}}{z^2}(1 + O(e^{-\Re(z)}))$ and $\eta(z) = \frac{e^{-2\gamma}}{z}(1 + O(e^{-\Re(z)})))$. And it turns out that the integral of $\left|\frac{\exp(2k(\omega))}{\omega^2}\eta(z+\omega)\right|$ over $\Re(\omega) = -\rho$ is now in $O(\frac{e^{-\rho y}}{\Re(z)})$. □

Lemma 5 *Let $\mathcal{K}$ be an arbitrary compact neighborhood of 0. When $y \to \infty$ and uniformly for $z \in \mathcal{K}$:*

$$f(y,z) = \delta(z)y + 2\delta(z) + e^{-2k(z)} - 1 + O(e^{-\rho y})\ , \tag{32}$$

with $\delta(z) = \int_0^z e^{-2k(\theta)} d\theta$.

Proof: We use again the inverse Laplace transform:

$$f(y,z) = f(y) + r(y,z) + \frac{1}{2i\pi}\int_{-\rho-i\infty}^{-\rho+i\infty} \tilde{f}(\omega,z)e^{\omega y} d\omega\ . \tag{33}$$

We know that $f(y) + r(y,z) = (\delta - \eta(z))(y+2) + e^{2k(z)} - 1 + O(e^{-\rho y})$.

Furthermore, since $\tilde{f}(\omega,z) = \frac{e^{2k(\omega)}}{\omega^2}\int_0^z e^{-2k(\theta+\omega)}d\theta$ and uniformly $e^{2k(\theta+\omega)} = \frac{e^{-2\gamma}}{(\theta+\omega)^2}(1+O(\frac{e^{-\omega-\theta}}{(\theta+\omega)}))$ and $\frac{e^{2k(\omega)}}{\omega^2} = e^{2\gamma}(1 + O(\frac{e^{-\omega}}{\omega}))$, therefore is uniformly when ω is large. Therfore $\tilde{f}(\omega,z)$ is uniformly absolutely integrable on $\Re(\omega) = -\rho$ and therefore

$$\int_{-\rho-i\infty}^{-\rho+i\infty} \tilde{f}(\omega,z)e^{\omega y} d\omega = O(e^{-\rho y})\ . \tag{34}$$

. □

5.2 Generalized DePoissonization with varying parameter y

Our aim is to compute the limit of $f_n(y_n)$ when both n and y tend to infinity. More rigorously, we assume that there is a sequence y_n and we assume that $\lim_{n\to\infty} \frac{n}{y_n} = \nu$ and we consider the three cases $\nu = 0$, $\nu = \infty$ and $0 < \nu < \infty$

First we consider the case $\nu = 0$ and we extend it to the case where n is fixed and $y \to \infty$. As we hinted in the introduction we expect that $\lim_{n\to\infty} f_n(y) = n$.

Theorem 3 (Asymptotics Case 1) *Assume that $y_n \to \infty$ and $\frac{n}{y_n} \to 0$. We have :*

$$f_n(y_n) = n - \frac{n^2 - n}{y_n} + \frac{1}{2}\frac{n^2 - n}{y_n^2} + O(\frac{n^3}{y_n^2}) \tag{35}$$

and the expansion can be continued up to any order.

Proof: We first work on the case $n \to \infty$. We remember that $f_n(y_n)$ is the sequence obtained from the "depoissonization" of the function $f(y_n, \frac{z}{y_n})$. We use the "diagonal depoissonization" lemma of [7], page 20. Since we assume that $z = \frac{n}{y}$ remains in a compact neighborhood $\mathcal{K}$ and we use lemma 5 to set the fact that in this case $f(y,z)$ remains bounded. This allows the depoissonization condition (O): $\forall \frac{z}{y_n} \in \mathcal{K} - \mathcal{C}$: $|f(y_n, \frac{z}{y_n})e^z|$ is bounded (therefore we get condition (O) with $\alpha = 0$). We have also condition (I):$\forall \frac{z}{y_n} \in \mathcal{K} \cap \mathcal{C}$: $f(y_n, z)$ is bounded. This implies via dePoissonization lemma:

$$f_n(y_n) = f(y, \frac{n}{y_n}) + O(\frac{1}{n}) \,. \tag{36}$$

In fact we can be more precise. We have

$$\delta(\frac{z}{y_n})(y_n+2) - e^{-2k(z/y_n)} = z - \frac{z^2}{y_n^2}(y_n - \frac{1}{2}) + O(\frac{z^3}{y_n^2}) \,. \tag{37}$$

Since the Poisson transform of the sequence $f_n(y_n) - n + \frac{n(n-1)}{y_n^2}(y_n - \frac{1}{2})$ is exactly $f(y_n, \frac{z}{y_n}) - z + \frac{z^2}{y_n^2}(y_n - \frac{1}{2})$ and by applying the dePoissonization lemma to $f(y_n, \frac{z}{y_n}) - z + \frac{z^2}{y_n^2}(y_n - \frac{1}{2})$ which is $O(\frac{z^3}{y_n^2})$. In this case we have the estimate:

$$f_n(y_n) = n - \frac{n^2-n}{y_n} + \frac{1}{2}\frac{n^2-n}{y_n^2} + O(\frac{n^3}{y_n^2}) \,. \tag{38}$$

The expansion of $\delta(\frac{z}{y_n})$ and $e^{-2k(z/y_n)}$ can be continued up to any order, converting each $\frac{z^k}{y_n^k}$ into $\frac{n(n-1)\cdots(n-k+1)}{y_n^k}$ to get the expansion of $f_n(y_n)$. □

Corollary 2 *Assume k fixed and $y \to \infty$:*

$$f_k(y_n) = k - \frac{k^2-k}{y} + \frac{1}{2}\frac{k^2-n}{y^2} + O(\frac{k^3}{y^2}) \tag{39}$$

and the expansion can be continued up to any order.

Proof: Since the dePoissonization error term is uniform and is the same as in the previous theorem but expressed in k and y it also converge to zero. □

Theorem 4 (Asymptotics case 2) *In the case $\nu = \infty$, we have:*

$$f_n(y_n) = \delta y + (2\delta - 1) - \frac{e^{-2\gamma}}{n+1}(2y_n+1) - \frac{e^{-2\gamma}}{(n+2)(n+1)} + O(e^{-\rho y_n}) + O(e^{-n}) \,. \tag{40}$$

Proof: For diagonal dePoissonization condition (O) we use lemma 1 to state that

$$\forall z \notin \mathcal{C}\ |z| = n:\ |f(y_n, z)e^{y_n z}| \leq P_{\lceil y_n \rceil}(n) = O((2\mu n)^{y_n}) \tag{41}$$

We use lemma 3 to get dePoissonization condition (I): $\forall z \in \mathcal{C}$: $|z| - n \in [-n\varepsilon, n\varepsilon] \Rightarrow |f(y_n, z) - f(y_n)| = O(\frac{y_n}{n})$. Thus we have the diagonal dePoissonization result:

$$f_n(y_n) - f(y_n) = f(y_n, \frac{n}{y_n}) - f(y_n) + O(\frac{y_n}{n^2}) \ . \tag{42}$$

We know from [7] that we can expand $f_n(y_n)$ to any order with the derivatives of $f(y_n, z)$ with respect to variable z at $z = \frac{n}{y_n}$. However we can indirectly get the expansion by noticing from lemma 4 that

$$f(y_n, \frac{n}{y_n}) = f(y_n) - (y_n + 2)y_n \frac{e^{-2\gamma}}{n} + y_n^2 \frac{e^{-2\gamma}}{n^2} + O(e^{-\rho y_n}) + O(e^{-n}) \ .$$

We notice that the Poisson transforms of the sequence $\frac{1}{n+1}$ and of the sequence $\frac{1}{(n+1)(n+2)}$ are respectively $\frac{1-e^{-z}}{z}$ and $\frac{1-e^{-z}-ze^{-z}}{z^2}$. Thus the Poisson transform of the sequence $f_n(y_n) - f(y_n) + (y_n + 2)y\frac{e^{-2\gamma}}{n+1} - y_n^2\frac{e^{-2\gamma}}{(n+1)(n+2)}$ is in $O(e^{-\rho y_n}) + O(e^{-n})$. Applying again the dePoissonization tool on these remaining terms we get that $f_n(y_n) - f(y_n) + (y_n + 2)y\frac{e^{-2\gamma}}{n+1} - y_n^2\frac{e^{-2\gamma}}{(n+1)(n+2)}$ is in $O(e^{-\rho y_n}) + O(e^{-n})$. $\square$

Theorem 5 (Asymptotics case 3) *In the case $0 < \nu < \infty$, we have for all interger k:*

$$f_n(y_n) = \delta y_n + (2\delta - 1) - \frac{e^{-2\gamma}}{n+1}(2y_n + 1) - \frac{e^{-2\gamma}}{(n+2)(n+1)} + O(e^{-\rho y_n}) + O(e^{-n}) \ . \tag{43}$$

Proof: It is the same proof as for the previous theorem with the notable exception that we cannot depoissonization condition (O) via lemma 1. Indeed we would get $|F(y_n, \frac{z}{y_n})| \leq (2\mu\frac{n}{y_n})^{y_n}$ which grow much faster than e^n. Since $\frac{z}{y_n}$ stays in a compact neighborhood we can use the lemma 5, namely that $\Re(z) < 0$ implies that $|f(y_n, \frac{z}{y_n})e^z| \leq |f(y_n, \frac{z}{y_n})| = O(1)$ in order to get the first depoissonization condition. $\square$

6 Conclusion

The linear bucket tree is a specific instanciation of geometric bucket tree for the dimension 1. Despite the apparent complexity of the basic equations, the Laplace transform has an explicit form. This allow to evaluate the average number of buckets on a segment of length y with n points, in most of the asymptotic regimes one can imagine for these models. We expect that the analysis can be extended to many other parameters such as: the average bucket depth, the average point depth, the average external path length, the variance of the number of buckets, *etc*. However the extension of this analysis to larger dimension remains an open problem. The main difficulty is in the fact that a bucket do not split the space in separate part, and therefore no divide and conquer based equation seems to apply.

References

[1] A. Rényi. On a one-dimensional random space-filling problem. Publ. Math. Res. Inst. Hung. Acad. Sci. 3, 109–127, 1958.

[2] G. Louchard. Exact and asymptotic distributions in digital and binary search trees. RAIRO Inform. Theor. Appl., 21(4):479–495, 1987

[3] P. Jacquet and M. Régnier. Normal limiting distribution of the size of tries. In PerformanceÕ87 (Brussels, 1987), pages 209–223. North-Holland, Amsterdam, 1988.

[4] L. Devroye. A study of trie-like structures under the density model. Ann. Appl. Probab., 2(2):402–434, 1992.

[5] P. Flajolet and B. Richmond. Generalized digital trees and their difference-differential equations. Random Structures Algorithms, 3(3):305–320, 1992.

[6] P. Kirschenhofer, H. Prodinger, andW. Szpankowski. Digital search trees again revisited: the internal path length perspective. SIAM J. Comput., 23(3):598–616, 1994.

[7] P. Jacquet and W. Szpankowski. Analytical de-Poissonization and its applications. Theoret. Comput. Sci., 201(1-2):1–62, 1998.

[8] M. Penrose. Random Parking, Sequential Adsorption and the Jamming Limit, Communications in Mathematical Physics, Springer, 2001.

[9] R. Gomes, H. F. Levison, K. Tsiganis, A. Morbidelli. Origin of the cataclysmic Late Heavy Bombardment period of the terrestrial planets, Nature 435, 2005.

[10] Zacharovas, Vytas; Hwang, Hsien-Kuei, A Charlier-Parseval approach to Poisson approximation and its applications, 2008 arXiv0810.4756Z

[11] F. Hubalek, H.-K. Hwang, W. Lew, H. Mahmoud, and H. Prodinger. A multivariate view of random bucket digital search trees. J. Algorithms, 44(1):121–158, 2002.