Binary Search Tree and 1-dimensional Random Packing

Yoshiaki Itoh

Institute of Statistical Mathematics, Tokyo, Japan

September 22, 1997

[summary by Hosam M. Mahmoud]

This talk surveys some models of random trees underlying continuous partitioning processes:

1. One-dimensional random sequential packing;
2. Kakutani's interval splitting;
3. The random sequential bisection model;
4. The continuous binary search tree.

1. One-Dimensional Random Sequential Packing

The first model was introduced in [5]. In this model we place a unit interval I_1 at a random position in the interval $[0, x]$. We assume that the initial point ξ_1 of the interval I_1 is Uniform-$[0, x-1]$. The interval (ξ_1, ξ_1+1) is removed and whichever among the remaining intervals $[0, \xi_1]$ and $[\xi_1+1, x]$ has length that permits further partitioning (i.e., greater than 1) is partitioned in a recursive fashion. The process continues as follows—if the intervals $I_1, I_2, \ldots, I_k$ have already been chosen, the next randomly chosen interval will be kept only if it does not intersect any of the intervals $I_1, I_2, \ldots, I_k$. In this case this interval will be denoted by I_{k+1}. If it intersects any of the intervals $I_1, I_2, \ldots, I_k$, we ignore it and choose a new interval. The procedure is continued until none of the lengths of gaps generated by the intervals placed in $[0, x]$ is greater than 1.

A parameter of interest is the number of intervals packed in $[0, x]$ by this procedure. We denote its mean value by $M(x)$. We can now formulate a differential equation (with delay) for $M(x)$. By conditioning on the initial point of the first interval and invoking its uniform distribution we obtain:

$$M(x+1) = \frac{2}{x}\int_0^x M(y)dy + 1;$$

for brevity many obvious boundary conditions are omitted in this overview of the talk. We can obtain the limiting behavior of this mean value via the Laplace transform and the method of undetermined coefficients. It then follows from a Tauberian theorem that, as $x \to \infty$,

$$\begin{aligned}\frac{M(x)}{x} &\to \int_0^\infty \exp\left(-2\int_0^t \frac{1-e^{-u}}{u}du\right)\, dt\\ &\doteq 0.748.\end{aligned}$$

Another parameter of interest is $L(x)$, the minimal gap length generated by the random packing. Again by conditioning on the position of the leftmost end of the first interval packed [2], we obtain an integral equation

$$P\left(L(x+1) \geq h\right) = \frac{1}{x}\int_0^x P\left(L(y) \geq h\right) P\left(L(x-y) \geq h\right)\, dy.$$

A similar integral equation can be obtained for the maximal gap length.

2. A Unified Model for Kakutani's Interval Splitting and Rényi's Random Packing

Rényi's partitioning process has an interpretation as a parking problem: One can park a car of length 1, if there is a space of length at least 1.

In a more general setting, one may consider parking cars (or packing intervals) of length ℓ, for a space of length at least 1. The expected number of cars is then

$$M(x+\ell) = \frac{1}{x}\int_0^x (M(y) + M(x-y) + 1)\, dy.$$

Rényi's problem is the case $\ell = 1$, whereas Kakutani considers the case $\ell = 0$. For this variation Komaki and Itoh [3] find the limiting behavior

$$\lim_{x\to\infty} \frac{M(x)}{x} = \int_0^\infty (1 + (1-\ell))\, e^{-(1-\ell)t} \exp\left(-2\int_0^t \frac{1-e^{-\ell u}}{u}\, du\right) dt.$$

For the probability distribution of the minimum of gaps, giving $f(x)$ for $0 \leq x \leq 1$, you get the functional form

$$f(x+\ell, h) = \frac{1}{x}\int_0^x f(x-y,h) f(y,h)\, dy.$$

3. The Height of a Continuous-Type Binary Search Tree

Consider a random permutation of $1, 2, \ldots, n$, with all $n!$ permutations equiprobable. Insert the elements of the permutation in a binary search tree. Let $H(n)$ be the height of the tree so obtained. This height satisfies the equation

$$P\left(H(n+1) \leq h\right) = \frac{1}{n}\sum_{m=0}^{n} P\left(H(n-m) \leq h-1\right) P\left(H(m) \leq h-1\right).$$

Note that the continuous analogue

$$P\left(H(x+1) \leq h\right) = \frac{1}{x}\int_0^x P\left(H(x-y) \leq h-1\right) P\left(H(y) \leq h-1\right) dy$$

is of the type we obtained in the continuous models considered earlier. Robson [6], Flajolet and Odlyzko [1], Mahmoud and Pittel [4] have considered heights of similar discrete-type random trees.

4. Random Sequential Bisection Model

Applying the idea for the analysis of random packing, a continuous model is studied as a random sequential bisection model [7].

Among the possible 2^d nodes at the d-th level, $1 \leq d$, of the associated tree the proportions of the expected number of the internal and external nodes are the Poisson-like expressions

$$\frac{1}{x}\sum_{k=d}^{\infty} \frac{(\log x)^k}{k!},$$

and

$$\frac{1}{x}\frac{(\log x)^{d-1}}{(d-1)!},$$

respectively.

Let $N_i(x,d)$ and $N_e(x,d)$ denote the numbers of the internal and external nodes at the d-th level respectively. Let $m_i(x,d)$ and $m_e(x,d)$ denote their expected values respectively. Then

$$m_i(x,d) = \frac{1}{x}\int_0^x (m_i(x-y,d-1) + m_i(y,d-1))\,dy.$$

From this we have

$$m_i(x,d) = \frac{2^d}{x}\sum_{k=d}^{\infty}\frac{(\log x)^k}{k!}, \qquad \text{for} \quad 1 \le x,$$

for $d = 0,1,2,\ldots$.

In any binary tree $N_i(x,d-1)$ internal nodes have $2N_i(x,d-1)$ son nodes, among which $N_i(x,d)$ are internal, therefore $N_e(x,d) = 2N_i(x,d-1) - N_i(x,d)$, for $d = 1,2,\ldots$. The expectation of this equality shows that for $d = 1,2,\ldots$, $m_e(x,d) = 2m_i(x,d-1) - m_i(x,d)$. As x and d increase to infinity in such a way that $d = c\log x$, we find

$$m_i(x,d) = \frac{1}{\sqrt{2\pi d}}e^{-d\gamma(c)} + O(1/d),$$

if $c > 1$. If $c < 1$,

$$m_i(x,d) = 2^d - \frac{1}{\sqrt{2\pi d}}e^{-d\gamma(c)} + O(1/d),$$

where $\gamma(c) = 1/c + \log(c/2) - 1$. It follows that

$$\lim_{x\to\infty} m_e(x,d) = \lim_{x\to\infty} m_i(x,d) = \begin{cases} 0, & \text{for } \hat{c} \le c < \infty; \\ \infty, & \text{for } 1 \le c < \hat{c}, \end{cases}$$

and, on the other hand,

$$\lim_{x\to\infty} m_e(x,d) = \lim_{x\to\infty} \{2^d - m_i(x,d)\} = \begin{cases} 0, & \text{for } 0 \le c < \check{c}; \\ \infty, & \text{for } \check{c} \le c < 1, \end{cases}$$

where $\hat{c} \doteq 4.311$ and $\check{c} \doteq 0.3734$ are the positive solutions of $\gamma(c) = 1/c + \log(c/2) - 1 = 0$.

Bibliography

[1] Flajolet (P.) and Odlyzko (A.). – The average height of binary trees and other simple trees. *Journal of Computer and System Sciences*, vol. 25, 1982, pp. 171–213.

[2] Itoh (Yoshiaki). – On the minimum of gaps generated by one-dimensional random packing. *Journal of Applied Probability*, vol. 17, n° 1, 1980, pp. 134–144.

[3] Komaki (Fumiyasu) and Itoh (Yoshiaki). – A unified model for Kakutani's interval splitting and Rényi's random packing. *Advances in Applied Probability*, vol. 24, n° 2, 1992, pp. 502–505.

[4] Mahmoud (Hosam) and Pittel (Boris). – On the most probable shape of a search tree grown from a random permutation. *SIAM Journal on Algebraic and Discrete Methods*, vol. 5, n° 1, 1984, pp. 69–81.

[5] Rényi (Alfréd). – On a one-dimensional problem concerning random space filling. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, vol. 3, n° 1/2, 1958, pp. 109–127.

[6] Robson (J. M.). – The height of binary search trees. *The Australian Computer Journal*, vol. 11, n° 4, 1979, pp. 151–153.

[7] Sibuya (Masaaki) and Itoh (Yoshiaki). – Random sequential bisection and its associated binary tree. *Annals of the Institute of Statistical Mathematics*, vol. 39, n° 1, 1987, pp. 69–84.