HEAVY TAILS IN ENGINEERING SYSTEMS

Ankit Bandyopadhyay

Table of contents

Highly Optimized Tolerance: Power Laws in Designed Systems

Introduction

Characteristic features of HOT systems include:

1. High efficiency, performance and robustness to designed-for uncertainties.
2. Hypersensitivity to design flaws and unanticipated perturbations.
3. Non-generic, specialized, structured configurations.
4. Power laws

The Model

- d-dimensional space denoted by X.
- Some knowledge of spatial distribution of probabilities of *initiating events*.
- Some *resource* to limit the size of events and some *constraint* on the amount of the resource.
- An *economic gain* on limiting the size of events.

The Model

An example -

Figure 1: Forest, Spark and Firebreak

Mathematical Model

- $p(\mathbf{x})$ - probability distribution for initiating events $\forall \mathbf{x} \in X$
- $A(\mathbf{x})$ - size of region affected by the event initiated at $\mathbf{x}$
- $C(\mathbf{x})$ - cost which scales as $A^{\alpha}(\mathbf{x}),\ \alpha > 0$
- $R(\mathbf{x})$ - resource restricting the event
- Constraint on resource: $\int_X R(\mathbf{x}) d\mathbf{x} = \kappa$
- Relation between affected region and resource: $A(\mathbf{x}) = R^{-\beta}(\mathbf{x})$

Analysis

$$
\begin{aligned}
&\mathbb{E}(A^{\alpha}) = \int_X p(\mathbf{x}) R^{-\alpha\beta}(\mathbf{x}) d\mathbf{x} \\
\Rightarrow\ &\delta \int_X [p(\mathbf{x}) R^{-\alpha\beta}(\mathbf{x}) d\mathbf{x} - \lambda R(\mathbf{x})] d\mathbf{x} = 0 \\
\Rightarrow\ &p(\mathbf{x}) R^{-\alpha\beta-1} = \text{const} \\
\Rightarrow\ &p(\mathbf{x}) \sim R^{\alpha\beta+1}(\mathbf{x}) \sim A^{-(\alpha+1/\beta)}(\mathbf{x}) \sim A^{-\gamma}(\mathbf{x})
\end{aligned}
$$

One Dimensional Case

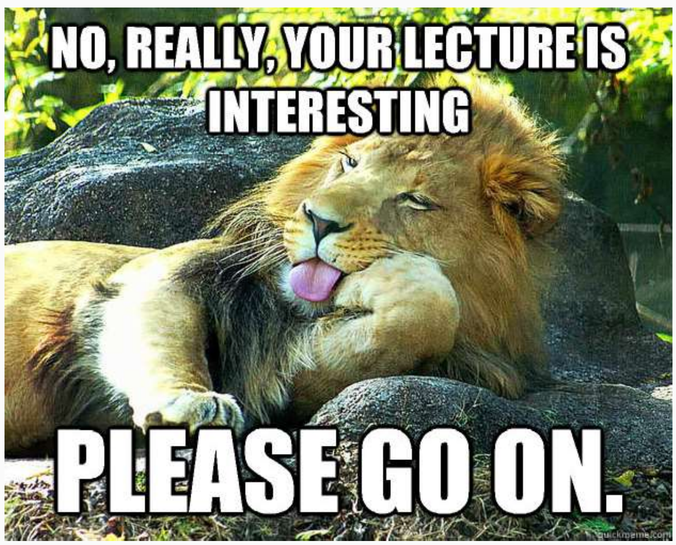
$p(x)$	$\bar{p}(x)$	$\bar{P}(A)$
$x^{-(q+1)}$	x^{-q}	$A^{-\gamma(1-1/q)}$
e^{-x}	e^{-x}	$A^{-\gamma}$
e^{-x^2}	$x^{-1}e^{-x^2}$	$A^{-\gamma}[log(A)]^{-1/2}$

Table 1: HOT states

Validity of the Model

- As we saw, $\bar{P}(A)$ is heavy-tailed for many common and light-tailed distributions.
- However, if $p(x)$ was distributed uniformly or if it was a Dirac-Delta, there would be no heavy-tails.

Go On...



Heuristically Optimized Trade-offs: A Paradigm for Power Laws in Internet

Introduction

A toy model of internet growth in which two objectives are optimized simultaneously -

- Connection Costs
- Transmission delays measured in hops

The Model

- Sequence of points $p_0, p_1, ..., p_n$ in the unit square, distributed uniformly at random.
- Sequence of undirected trees $T_0, T_1, ..., T_n$ with T_0 consisting of p_0 alone.
- Define h_i to be the number of hops from p_i to p_0.
- Define d_{ij} to be the Euclidean distance between i and j.
- Let α to be a fixed number but is a function of n.
- T_i is defined as T_{i-1} with the point i and edge $[i, j]$ added.
- $j < i$ minimizes $f_i(j) = \alpha d_{ij} + h_j$.
- Neighborhood $N_k(i) = j | [i, j] \in T_k$.
- Let $T = T_n$ and $N(i)$ be the neighborhood of i in T.

Result

1. If $\alpha < 1/\sqrt{2}$, then T is a star with p_0 at its center.
2. If $\alpha = \Omega(\sqrt{n})$, then the degree distribution of T is exponential.
   $E[|i : \text{ degree of } i \geq D|] < n^2 e^{-cD}$.
3. If $\alpha > 4$ and $\alpha = o(\sqrt{n})$, then degree distribution is heavy tailed.
   $E[|i : \text{ degree of } i \geq D|] > c(D/n)^{-\beta}$ for some constants β and c.
   Specifically, for $\alpha = o(n^{1/3})$, we have $\beta \geq 1/6$ and $c = O(\alpha^{-1/2})$.

Proof

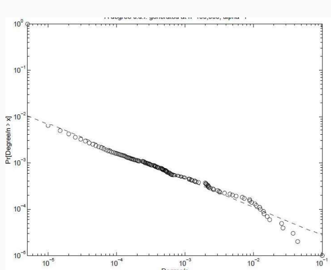
- Concentrate on points near P_0.
- Consider points $j \in N_0$, with $d_{0j} \leq 2\alpha$.
- Without loss of generality, we assume that p_0 is located at least $2/\alpha$ from the boundary.
- Will prove two lemmas for a point i which arrives so that $i \in N_0$ and $3/2\alpha > d_{i0} > 1/\alpha$. Let $r(i) = d_{i0} - 1/\alpha$.

Proof

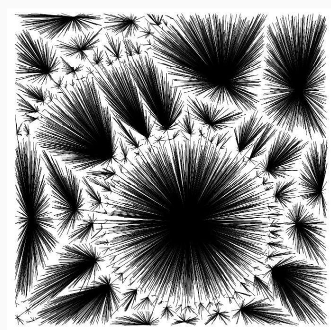
We will prove the following lemmas -

- Every point arriving after i inside the circle of radius $1/4\ r(i)$ around i will link to i.
- No point j will link to i unless $|\angle p_j p_0 p_i| \leq \sqrt{2.5\alpha r(i)}$ and $d_{j0} \geq 1/2 r(i) + 1/\alpha$.

Simulation Results



(a) CCDF for $\alpha = 4$ and $n = 100,000$



(b) Tree generated

Too Hot?

Informational Theory of the Statistical Structure of Language

Scope of the Model

This model is based upon the consideration of a single speaker and a single receiver.

Three elements were considered -

- The structure of language, or message.
- The way in which information is coded by the brain.
- The economical criterion of matching which links 1 & 2.

The Problem to be Solved

It is easiest to find out the language given the economical criterion and the way in which information is coded by the brain.

Why?

The Problem to be Solved

It is easiest to find out the language given the economical criterion and the way in which information is coded by the brain.

We would not require a preliminary numerical representation of the structure of language which is a very difficult and ambiguous task.

The Problem to be Solved

- We will assume language as a random sequence of concrete entities like words.
- Definition of a word is itself ambiguous but consider the word in its inflectional form and not lexical form.
- Empirically, power laws have been observed for inflectional forms but not lexical forms.
- Aim - nth word by order of decreasing frequency must be represented by the nth sequence of signs by order of increasing cost.

Cost of Words

Cost of the nth sequence can be written as -

$$C_n = \lceil \log_M(n + m) + j_0 \rceil$$

where M is the alphabet size, m and j_0 are some constants.

Maximizing Information per Unit Cost

The user sends a nth rank word with a probability p_n. Then the average information per word $H = -\sum_n p_n \log(p_n)$ and the average cost per word $C = \sum_n p_n C_n$. We need to choose p_n to minimize

$$A = \frac{C}{H} = \frac{\sum_n p_n C_n}{-\sum_n p_n \log_2(p_n)}$$

which gives

$$\frac{dA}{dp_n} = \frac{C_n H + C \log_2(e p_n)}{H^2} = 0$$

$$\Rightarrow \; p_n = 2^{-HC_n/C}/e$$

$$\Rightarrow \; p_n \sim n^{-H \log_M(2)/C}$$

Experimental Results

The author gives the following canonical law -

$$p_n = P(n+m)^{-B}$$

where P, M and B are positive constants. This equation was an improvement over Zipf's law -

$$p_n = Pn^{-1}$$

Experimental Results

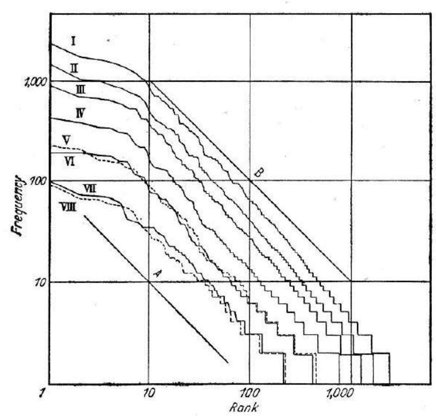


Figure 1. Rank-frequency distributions of Helen B. (with paranoid schizophrenia) in samples of (I) 50,000 words; (II) 30,000 words; (III) 20,000 words; (IV) 10,000 words; (V and VI) 5,000 words, and (VII and VIII) 2,000 words. (From *Arch. Neurol. Psychiat.* 49 (1943) 831)

Figure 3: Rank Frequency Distribution in Schizophrenics

Experimental Results

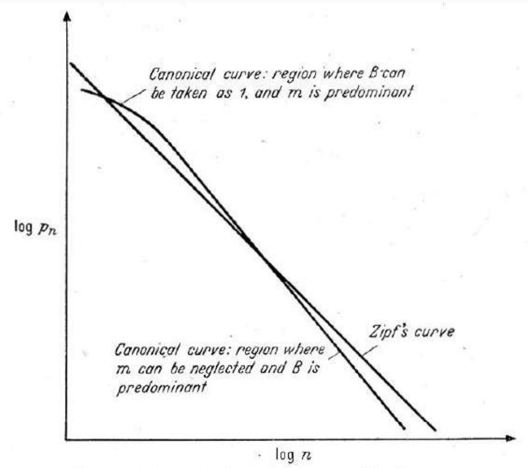


Figure 2. Canonical Rank-frequency distribution
$p_n = P(n + m)^{-B}$ as compared with Zipf's "law" $pn = P^{n-1}$

Figure 4: Rank Frequency Distribution in English

Experimental Results

Fun Fact - In one sample of words in the English language, the most frequently occurring word, "the", accounts for nearly 7% of all the words (69,971 out of slightly over 1 million). True to Zipf's Law, the second-place word "of" accounts for slightly over 3.5% of words (36,411 occurrences), followed by "and" (28,852). Only about 135 words are needed to account for half the sample of words in a large sample.

Conclusion

Conclusion

- We studied different models to explain how heavy tails emerge in optimized engineering systems.
- We looked at examples of *firebreaks*, *internet topology* and *language*.

It’s Over!

References

References

- Michael Mitzenmacher. A Brief History of Generative Models for Power Law and Lognormal Distributions
- J. M. Carlson and J. Doyle. Highly Optimized Tolerance: A Mechanism for Power Laws in Designed Systems.
- A. Fabrikant, E. Koutsoupias, and C. H. Papadimitriou. Heuristically Optimized Tradeoffs: A New Paradigm for Power Laws in the Internet.
- B. Mandelbrot. An Informational Theory of the Statistical Structure of Languages.