Benfordness of lower-dimensional measurements resulting from box fragmentation

Livia Betti, Zoe McDonald

SMALL REU Williams College, Probability and Analytic Number Theory
Joint work with Santiago Miguel Velazquez Iannuzzelli, Jack Miller, and Steven J. Miller

Young Mathematician's Conference
August 14th, 2022

Outline

1. Background
   - Benford's Law
   - Fragmentation Processes
2. Benford Behavior from Fragmentation Processes
   - Probability Background
   - Benfordness of Volumes resulting from Box Fragmentation
3. Lower-dimensional Volumes in Box Fragmentation
   - Uniform Distributions: Experimental Data
   - Proof for Perimeter
4. References

Outline

1. Background
   - Benford's Law
   - Fragmentation Processes
2. Benford Behavior from Fragmentation Processes
3. Lower-dimensional Volumes in Box Fragmentation
4. References

Benford's Law

Definition (Benford's Law)

The frequencies of first digits d (in base B) follow closely to the logarithmic relation:

$$\log_B\left(\frac{d+1}{d}\right)$$

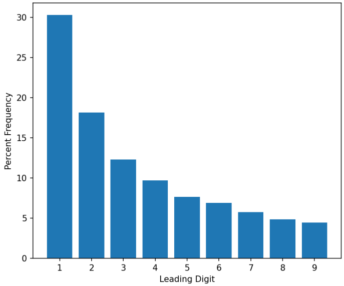
Benford's Law

Definition (Benford's Law)

The frequencies of first digits d (in base B) follow closely to the logarithmic relation:

$$\log_B\left(\frac{d+1}{d}\right)$$

Base 10:



How to show Benford Behavior

Definition (Significand)

Given a base $B \geq 2$, we write $x = aB^n$ (scientific notation) with $|a| \in [1, B)$, $n \in \mathbb{Z}$. The *significand* of x is $|a|$.

- Example ($B = 10$): $1022 = 1.022 \times 10^3$ has significand 1.022.

How to show Benford Behavior

Definition (Significand)

Given a base $B \geq 2$, we write $x = aB^n$ (scientific notation) with $|a| \in [1, B)$, $n \in \mathbb{Z}$. The *significand* of x is $|a|$.

- Example ($B = 10$): $1022 = 1.022 \times 10^3$ has significand 1.022.

Definition (Strong Benford's Law)

A data set satisfies *Strong Benford's Law base B* if the probability of observing a significand of at most s in base B is $\log_B s$

How to show Benford Behavior

Definition (Equidistributed)

A sequence $\{x_n\} \subset [0, 1]$ is equidistributed in $[0, 1]$ if for all $[a, b] \subset [0, 1]$,

$$\lim_{N\to\infty} \frac{\#\{n : |n| \leq N, x_n \in [a, b]\}}{2N + 1} = b - a$$

How to show Benford Behavior

Definition (Equidistributed)

A sequence $\{x_n\} \subset [0, 1]$ is equidistributed in $[0, 1]$ if for all $[a, b] \subset [0, 1]$,

$$\lim_{N\to\infty} \frac{\#\{n : |n| \leq N, x_n \in [a, b]\}}{2N + 1} = b - a$$

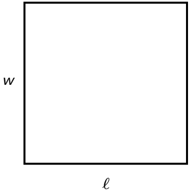
Theorem

If $y_n = \log_B x_n$ is equidistributed $\pmod 1$ then x_n is Benford base B.

Box Fragmentation

Fragmenting a rectangle in 2-dimensions:

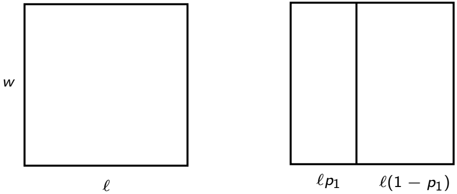
- Begin with an 2-dimensional box and density functions f_1, f_2 on $(0, 1)$.



Box Fragmentation

Fragmenting a rectangle in 2-dimensions:

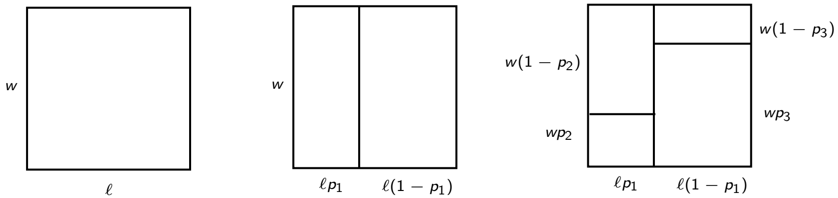
- Begin with an 2-dimensional box and density functions f_1, f_2 on $(0, 1)$.
- Draw a proportion p_1 from f_1 and cut the box along its horizontal axis at proportion p_1.



Box Fragmentation

Fragmenting a rectangle in 2-dimensions:

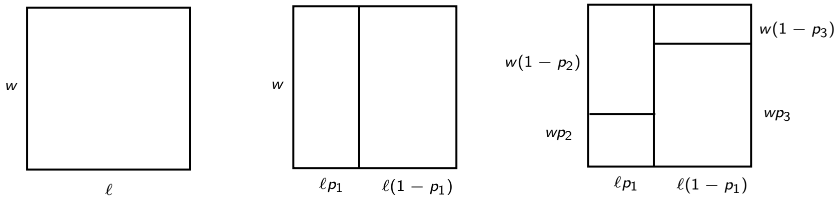
- Begin with an 2-dimensional box and density functions f_1, f_2 on $(0, 1)$.
- Draw a proportion p_1 from f_1 and cut the box along its horizontal axis at proportion p_1.
- Draw p_2 and p_3 from f_2. Cut one subrectangle at proportion p_2 along the vertical axis and the other subrectangle at proportion p_3 along the vertical axis.



Box Fragmentation

Fragmenting a rectangle in 2-dimensions:

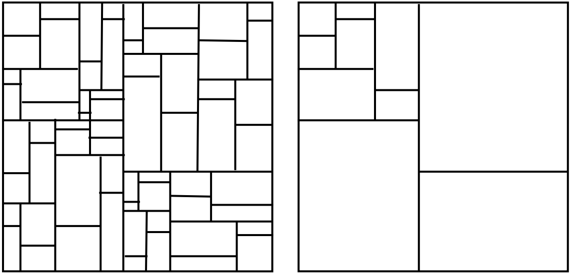
- Begin with an 2-dimensional box and density functions f_1, f_2 on $(0, 1)$.
- Draw a proportion p_1 from f_1 and cut the box along its horizontal axis at proportion p_1.
- Draw p_2 and p_3 from f_2. Cut one subrectangle at proportion p_2 along the vertical axis and the other subrectangle at proportion p_3 along the vertical axis.
- Continue this way for n iterations.



Decomposition Models

We can vary the way we apply the fragmentation to the box:

- Unrestricted Decomposition Model: Each piece decomposes.

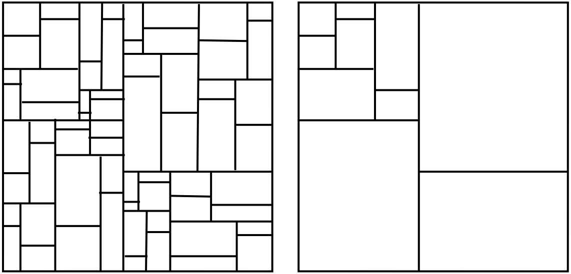


Unrestricted vs. Restricted Fragmentation

Decomposition Models

We can vary the way we apply the fragmentation to the box:

- Unrestricted Decomposition Model: Each piece decomposes.
- Restricted Decomposition Model: Only one resulting piece decomposes further.



Unrestricted vs. Restricted Fragmentation

Questions we can ask

What has been studied: Benfordness of m-dimensional volumes resulting from fragmentation of an m-dimensional box. For example:

- Area of 2-dimensional rectangle
- Volume of 3-dimensional box

These processes are known to exhibit Benford behavior as the number of iterations, n, goes to infinity!

Questions we can ask

What has been studied: Benfordness of m-dimensional volumes resulting from fragmentation of an m-dimensional box. For example:

- Area of 2-dimensional rectangle
- Volume of 3-dimensional box

These processes are known to exhibit Benford behavior as the number of iterations, n, goes to infinity!

What we are studying: Benfordness of lower-dimensional volumes resulting from hyper-box fragmentation. For example:

- Perimeter of 2-dimensional rectangle
- Surface Area of 3-dimensional box
- Frame of 3-dimensional box

Do these processes exhibit Benford behavior?

Questions we can ask

What has been studied: Benfordness of m-dimensional volumes resulting from fragmentation of an m-dimensional box. For example:

- Area of 2-dimensional rectangle
- Volume of 3-dimensional box

These processes are known to exhibit Benford behavior as the number of iterations, n, goes to infinity!

What we are studying: Benfordness of lower-dimensional volumes resulting from hyper-box fragmentation. For example:

- Perimeter of 2-dimensional rectangle
- Surface Area of 3-dimensional box
- Frame of 3-dimensional box

Do these processes exhibit Benford behavior? Yes!

Outline

1 Background

2 Benford Behavior from Fragmentation Processes
- Probability Background
- Benfordness of Volumes resulting from Box Fragmentation

3 Lower-dimensional Volumes in Box Fragmentation

4 References

Probability

Definition (Probability Density Function (pdf))

Let X be a continuous random variable. If f satisfies

$$\mathbb{P}(X \leq b) = \int_{-\infty}^{b} f(t)dt$$

then f is the *probability density function* of X.

Probability

Definition (Probability Density Function (pdf))

Let X be a continuous random variable. If f satisfies

$$\mathbb{P}(X \leq b) = \int_{-\infty}^{b} f(t)dt$$

then f is the *probability density function* of X.

Definition (Convolution)

If X, Y are independent continuous random variables with pdfs f_X, f_Y, then the pdf of $X + Y$ is the called the *convolution* of f_X and f_Y, and

$$f_{X+Y}(z) = (f_X * f_Y)(z) = \int_{-\infty}^{\infty} f_X(t) f_Y(z-t)dt$$

Mellin Convolution and Mellin Transform

Definition (Mellin Convolution)

If X, Y are independent continuous random variables with pdfs f_X, f_Y, then the pdf of $X \cdot Y$ is the called the *Mellin convolution* of f_X and f_Y, and

$$f_{X \cdot Y}(z) = (f_X *_{\mathcal{M}} f_Y)(z) = \int_{-\infty}^{\infty} f_X(t) f_Y\left(\frac{z}{t}\right) \frac{dt}{t}$$

Mellin Transform

Definition (Mellin Transform)

Let f be a continuous real-valued function on $[0, \infty)$. The *Mellin Transform* of f, denoted $\mathcal{M}_f$, is

$$\mathcal{M}_f(s) = \int_0^\infty f(t)t^{s-1}dt$$

Mellin Transform

Definition (Mellin Transform)

Let f be a continuous real-valued function on $[0, \infty)$. The *Mellin Transform* of f, denoted $\mathcal{M}_f$, is

$$\mathcal{M}_f(s) = \int_0^\infty f(t) t^{s-1} dt$$

Theorem

Let X, Y be independent continuous random variables with pdfs f_X, f_Y defined on $[0, \infty)$. Then,

$$\mathcal{M}_{(f_X *_{\mathcal{M}} f_Y)}(s) = \mathcal{M}_{f_X}(s) \cdot \mathcal{M}_{f_Y}(s)$$

Benfordness of Volumes resulting from Box Fragmentation

From SMALL 2018 and Irfan Durmić:

Consider an m-dimensional box with volume V. For random variables $X_1, \dots, X_N$, fix a continuous probability density f_k on $(0,1)$ such that

$$\lim_{N\to\infty} \sum_{\substack{\ell=-\infty \\ \ell\neq 0}}^{\infty} \left| \prod_{k=1}^{Nm} \mathcal{M}_{f_k}\left(1 - \frac{2\pi i \ell}{\log 10}\right) \right| = 0.$$

Benfordness of Volumes resulting from Box Fragmentation

From SMALL 2018 and Irfan Durmić:

Consider an m-dimensional box with volume V. For random variables $X_1, \dots, X_N$, fix a continuous probability density f_k on $(0,1)$ such that

$$\lim_{N\to\infty} \sum_{\substack{\ell=-\infty \\ \ell\neq 0}}^{\infty} \left| \prod_{k=1}^{Nm} \mathcal{M}_{f_k}\left(1 - \frac{2\pi i \ell}{\log 10}\right) \right| = 0.$$

Then, if unrestricted fragmentation is performed on the m-dimensional box, the resulting m-dimensional volumes follow Benford behavior, and

$$\begin{aligned} &|\mathbb{P}(\log(X_1 \cdots X_N) \pmod 1) \in [a,b] - (b-a))| \\ &\leq (b-a) \cdot \sum_{\substack{\ell=-\infty \\ \ell\neq 0}}^{\infty} \left| \prod_{k=1}^{Nm} \mathcal{M}_{f_k}\left(1 - \frac{2\pi i \ell}{\log 10}\right) \right| \end{aligned}$$

Outline

1 Background

2 Benford Behavior from Fragmentation Processes

3 Lower-dimensional Volumes in Box Fragmentation
- Uniform Distributions: Experimental Data
- Proof for Perimeter

4 References

Benfordness of Volumes

From before, we have

$$\lim_{N\to\infty} \sum_{\substack{\ell=-\infty \\ \ell\neq 0}}^{\infty} \left| \prod_{k=1}^{Nm} \mathcal{M}_{f_k}\left(1 - \frac{2\pi i \ell}{\log 10}\right) \right| = 0.$$

Benfordness of Volumes

From before, we have

$$\lim_{N\to\infty} \sum_{\substack{\ell=-\infty \\ \ell\neq 0}}^{\infty} \left| \prod_{k=1}^{Nm} \mathcal{M}_{f_k}\left(1 - \frac{2\pi i \ell}{\log 10}\right) \right| = 0.$$

Where does this condition break when we look at lower-dimensional volumes?

Benfordness of Volumes

From before, we have

$$\lim_{N\to\infty} \sum_{\substack{\ell=-\infty \\ \ell\neq 0}}^{\infty} \left| \prod_{k=1}^{Nm} \mathcal{M}_{f_k}\left(1 - \frac{2\pi i\ell}{\log 10}\right) \right| = 0.$$

Where does this condition break when we look at lower-dimensional volumes?

Mellin Transforms do not behave nicely under addition! For $X \cdot Y$,

$$\mathcal{M}_{(f_X *_{\mathcal{M}} f_Y)}(s) = \mathcal{M}_{f_X}(s) \cdot \mathcal{M}_{f_Y}(s)$$

But, for $X + Y$

$$\mathcal{M}_{(f_X * f_Y)}(s) \neq \mathcal{M}_{f_X}(s) \cdot \mathcal{M}_{f_Y}(s)$$

Perimeter

Let's look at the fragmentation process on a 2-dimensional rectangle with dimensions ℓ, w:

Perimeter

Let's look at the fragmentation process on a 2-dimensional rectangle with dimensions ℓ, w:

- Draw proportions p_i and $q_{2i+1}, q_{2(i+1)}$ from distributions with a piecewise continuous pdf.

Perimeter

Let's look at the fragmentation process on a 2-dimensional rectangle with dimensions ℓ, w:

- Draw proportions p_i and $q_{2i+1}, q_{2(i+1)}$ from distributions with a piecewise continuous pdf.
- The perimeter of an arbitrary box after N iterations is

$$Per = 2\left(\underbrace{\ell\prod_{i=1}^{k} p_i \prod_{j=k+1}^{N}(1-p_j)} + \underbrace{w\prod_{i=1}^{k} q_i \prod_{j=k+1}^{N}(1-q_j)}\right)$$

Uniform Distributions

First, we looked at the fragmentation process resulting from p_i and q_{2i}, q_{2i+1} being drawn from uniform distributions.

Lemma

The product of n uniform distributions has pdf

$$f_n(x) = \begin{cases} \frac{(-\ln(x))^{n-1}}{(n-1)!} & \textit{if } x \in (0,1] \\ 0 & \textit{otherwise} \end{cases}$$

Integration Attempt

Proportion of perimeters with leading digit d is given by:

$$\sum_m \mathbb{P}(d \cdot 10^{-m} \leq Per < (d+1)10^{-m})$$

$$= \sum_m \int_0^{d \cdot 10^{-m}} f_n(y) \int_{d \cdot 10^{-m}}^{(d+1) \cdot 10^{-m}} f_n(x) dx dy + \int_{d \cdot 10^{-m}}^{(d+1) \cdot 10^{-m}} f_n(y) \int_0^{(d+1) \cdot 10^{-m}} f_n(x) dx dy$$

Approximation

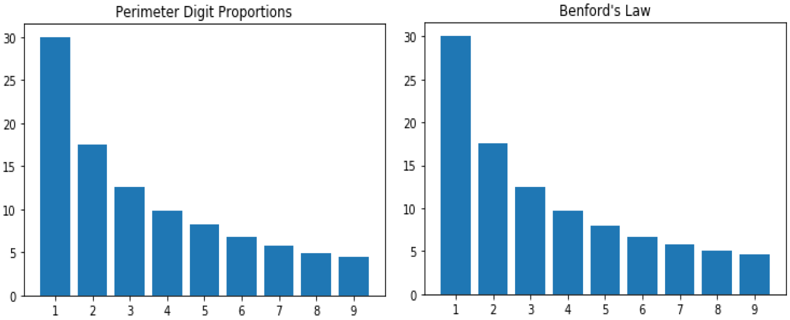
For 8 iterations of fragmentation, we can approximate the proportion of perimeters with leading digit 1:

$$\sum_{m=0}^{100}\left(\int_0^{10^{-m}} f_8(y)\int_{10^{-m}-y}^{2\cdot 10^{-m}-y} f_8(x)dxdy + \int_{10^{-m}}^{2\cdot 10^{-m}} f_8(y)\int_0^{2\cdot 10^{-m}-y} f_8(x)dxdy\right)$$

$$\approx \quad 0.301032\dots$$

$$\approx \quad \log_{10}\left(\frac{1+1}{1}\right) = 0.3010299\dots$$

Approximation



Modeling 8 iterations of the Unrestricted Model.

Using the Maximum of two Benford RVs

Let $L(\cdot, B) : \mathbb{R} \to \mathbb{N}$ denote the function that returns the leading digit of a real number base B. (Note this is not the same as the significand.)

When is $L(X_n + Y_n, B) = L(\max\{X_n, Y_n\}, B)$?

Using the Maximum of two Benford RVs

When is $L(X_n + Y_n, B) = L(\max\{X_n, Y_n\}, B)$?

Theorem (SMALL 2022)

Let f_{X_n} and f_{Y_n} be pdfs of X_n, Y_n respectively. Suppose that f_{X_n}, f_{Y_n} are piecewise continuous pdfs. Then, for any ε, the probability that $L(X_n + Y_n, B) = L(\max\{X_n, Y_n\}, B)$ is at least $(1-\varepsilon)\left(1 - O\left(\frac{1}{\sqrt{n}}\right)\right)$.

Using the Maximum of two Benford RVs

When is $L(X_n + Y_n, B) = L(\max\{X_n, Y_n\}, B)$?

Theorem (SMALL 2022)

Let f_{X_n} and f_{Y_n} be pdfs of X_n, Y_n respectively. Suppose that f_{X_n}, f_{Y_n} are piecewise continuous pdfs. Then, for any ε, the probability that $L(X_n + Y_n, B) = L(\max\{X_n, Y_n\}, B)$ is at least $(1-\varepsilon)\left(1 - O\left(\frac{1}{\sqrt{n}}\right)\right)$.

Theorem (SMALL 2022)

$\max\{X_n, Y_n\}$ *converges to Benford behavior base B as $n \to \infty$.*

Using the Maximum of two Benford RV's

We have

$$\begin{aligned}\mathbb{P}(L(X_n+Y_n)=d) &= \left(1-O\left(\frac{1}{\sqrt{n}}\right)\right)(1-\varepsilon)\mathbb{P}(L(\max\{X_n, Y_n\})=d) \\ &+ \left(1-\left(1-O\left(\frac{1}{\sqrt{n}}\right)\right)(1-\varepsilon)\right)\mathbb{P}^*(L(X_n+Y_n)=d)\end{aligned}$$

Using the Maximum of two Benford RV's

We have

$$\begin{aligned}\mathbb{P}(L(X_n+Y_n)=d) &= \left(1-O\left(\frac{1}{\sqrt{n}}\right)\right)(1-\varepsilon)\mathbb{P}(L(\max\{X_n,Y_n\})=d)\\ &+ \left(1-\left(1-O\left(\frac{1}{\sqrt{n}}\right)\right)(1-\varepsilon)\right)\mathbb{P}^*(L(X_n+Y_n)=d)\end{aligned}$$

Substitute in

$$\begin{aligned}\mathbb{P}(L(\max\{X_n,Y_n\})=d) &= \mathbb{P}(X_n\geq Y_n)\cdot\mathbb{P}(L(X_n)=d))\\ &+ \mathbb{P}(Y_n\geq X_n)\cdot\mathbb{P}(L(Y_n)=d))\end{aligned}$$

Using the Maximum of two Benford RV's

We have

$$\begin{aligned}\mathbb{P}(L(X_n+Y_n)=d) &= \left(1-O\left(\frac{1}{\sqrt{n}}\right)\right)(1-\varepsilon)\mathbb{P}(L(\max\{X_n,Y_n\})=d)\\ &+ \left(1-\left(1-O\left(\frac{1}{\sqrt{n}}\right)\right)(1-\varepsilon)\right)\mathbb{P}^*(L(X_n+Y_n)=d)\end{aligned}$$

Substitute in

$$\begin{aligned}\mathbb{P}(L(\max\{X_n,Y_n\})=d) &= \mathbb{P}(X_n\geq Y_n)\cdot\mathbb{P}(L(X_n)=d)\\ &+ \mathbb{P}(Y_n\geq X_n)\cdot\mathbb{P}(L(Y_n)=d)\end{aligned}$$

Using the Maximum of two Benford RV's

Cancelling out terms and using that

$$\mathbb{P}(L(X_n) = d) = \log_B\left(\frac{d+1}{d}\right) \quad \mathbb{P}(L(Y_n) = d) = \log_B\left(\frac{d+1}{d}\right),$$

we get

$$\mathbb{P}(L(X_n + Y_n) = d) = \log_B\left(\frac{d+1}{d}\right)$$

Using the Maximum of two Benford RV's

Cancelling out terms and using that

$$\mathbb{P}(L(X_n) = d) = \log_B\left(\frac{d+1}{d}\right) \qquad \mathbb{P}(L(Y_n) = d) = \log_B\left(\frac{d+1}{d}\right),$$

we get

$$\mathbb{P}(L(X_n + Y_n) = d) = \log_B\left(\frac{d+1}{d}\right)$$

Theorem (SMALL 2022)

Let $k \leq m$. The k-dimensional volumes of boxes resulting from unrestricted fragmentation of an m-dimensional box with proportions drawn from piecewise continuous pdfs follow Benford behavior.

Using the Maximum of two Benford RV's

Cancelling out terms and using that

$$\mathbb{P}(L(X_n) = d) = \log_B\left(\frac{d+1}{d}\right) \quad \mathbb{P}(L(Y_n) = d) = \log_B\left(\frac{d+1}{d}\right),$$

we get

$$\mathbb{P}(L(X_n + Y_n) = d) = \log_B\left(\frac{d+1}{d}\right)$$

Theorem (SMALL 2022)

Let $k \leq m$. The k-dimensional volumes of boxes resulting from unrestricted fragmentation of an m-dimensional box with proportions drawn from piecewise continuous pdfs follow Benford behavior.

Note: This is not Strong Benford Behavior!

Future Work

- Prove Strong Benford Behavior for the lower-dimensional volumes resulting from fragmentation processes
- Determine classes of functions which satisfy the conditions necessary for Benford behavior and generalize our proof to other distributions
- Expand the library of distributions that satisfy the conditions of the Mellin Transform theorem used to prove the original volume problem

Outline

1 Background

2 Benford Behavior from Fragmentation Processes

3 Lower-dimensional Volumes in Box Fragmentation

4 References

Acknowledgments

Thank you!

This research was conducted as part of the **2022 SMALL REU program** at Williams College, and was funded by **NSF Grant DMS1947438**, **Harvey Mudd College**, and **Williams College** funds. We thank Steven J. Miller and our colleagues from the 2022 SMALL REU program for many helpful conversations.

Reference

- Thealexa Becker, David Burt, Taylor C. Corcoran, Alec Greaves-Tunnell, Joseph R. Iafrate, Joy Jing, Steven J. Miller, Jaclyn D. Porfilio, Ryan Ronan, Jirapat Samranvedhya, Frederick W. Strauch, and Blaine Talbut. *Benford's law and continuous dependent random variables*. Annals of Physics vol. 388, 2018, pp. 350-381.
- Durmić, Irfan. *Benford Behavior of a Higher-Dimensional Fragmentation Process*. 2022.
- Jang, Dennis et al. *Chains of Distributions, Hierarchical Bayesian Models and Benford's Law*. 2008.