Semester Project Report
Infinite Divisibility of Information

Salim Najib - Supervised by Yanina Shkel and Anuj Yadav

EPFL - Spring 2024

1 Introduction

Physical quantities like space, time and mass can be divided infinitely: hours can be divided into minutes, then minutes into seconds, etc. A new quantity of interest was theorized during the XXth century: information, as studied by Shannon in his 1948 memorandum *A mathematical theory of comunication* [1]. Can it too be divided arbitrarily many times?

Cheuk Ting Li took it upon himself to answer this very question in his paper *Infinite Divisibility of Information* [2], motivated by Kolmogorov's constructions in probability theory as per the divisibility of *randomness* [3]. He finds that the answer is *no* in general, yet he proves fundamental bounds on the approximate divisibility of information. The main result, which will be detailed in 5.4, essentially states that a discrete random variable X can be divided into n i.i.d pieces $Z_1, \dots, Z_n$ from which X can be reconstructed and such that for all i

$$H(Z_i) = \frac{e}{e-1}\frac{H(X)}{n} + O\left(\frac{\log(n)}{\sqrt{n}}\right).$$

His work has important applications in areas such as secret distributed storage, where some piece of information is divided into many pieces, and all the pieces should be needed to reconstruct the initial piece of information.
Moreover, this content is closely related to work on entropy couplings ([4], [5], [6]) and the functional representation lemma ([5], [7]).

This report is a walkthrough and discussion of Professor Li's paper. The aim is to make its content more accessible and to comment on its results and their implications.

Contents

2 Mathematical background

This first section delves into some background notions and definitions that will be needed later on. It is an introduction to the tools we will be using later.

2.1 Notation and conventions

- log is taken to the base 2 when unspecified. The natural logarithm is ln.
- $h_2(p)$ is the binary entropy of p: $h_2(p) = -p\log(p) - (1-p)\log(1-p)$
- $[\![x, y]\!] = [x, y] \cap \mathbb{Z}$.
- When $f: A \to B$ is a function, $\mathrm{Im}(f) = \{f(x) \mid x \in A\}$ is the range or image of f.
- When X and Y are random variables, $X \sim Y \iff F_X = F_Y$.
- When $\{x_i\}_{i\in J}$ is a finite or infinite sequence, then $x^n = (x_1, x_2, \ldots, x_n)$.
- w.l.o.g = without loss of generality
- s.t = such that
- cdf = cumulative distribution function
- pdf = probability density function
- pmf = probability mass function
- RV = random variable
- When $X \sim F$ where F is a cdf, we may write $\mathbb{E}(F) = \mathbb{E}(X)$. Also, we define $F_X = F$ and the pmf of X is written $p_X(x) = \mathbb{P}(X = x)$.

More notations are defined in the remainder of this section.

2.2 Majorization in $\mathcal{S}_n$

Throughout the paper, comparing probability mass functions is of critical importance. There are several rudimentary ways to do that, two of which will be listed in this subsection, and one more in the next one.

2.2.1 Majorization in $\mathbb{R}^n$

Let $x, y \in \mathbb{R}^n$. x **majorizes** y i.e $y \preceq_m x$ if, for all $k \in [\![1, n]\!]$ we have

$$\sum_{i=1}^{k} x_{(i)} \geq \sum_{i=1}^{k} y_{(i)}$$

where $(v_{(i)})_{i=1}^n$ is a nonincreasing sorting of $v \in \mathbb{R}^n$. This defines a preorder in $\mathbb{R}^n$ (reflexive and transitive).

2.2.2 Equivalent definition in $\mathcal{S}_n$

Let $\mathcal{S}_n$ be the probability simplex in $\mathbb{R}^n$ with $n \in \mathbb{N} \cup \{+\infty\}$, i.e

$$\mathcal{S}_n = \left\{ p \in \mathbb{R}^n \mid p_i \geq 0,\ \sum_{i=1}^{n} p_i = 1 \right\}$$

Then let $p, q \in \mathcal{S}_n$. $q \preceq_m p$ is equivalent to

$$\forall k \in [\![1, n]\!] \quad \max_{B \subseteq [\![1,n]\!] : |B| \leq k} q(B) \leq \max_{A \subseteq [\![1,n]\!] : |A| \leq k} p(A)$$

where $p(A) = \sum_{a \in A} p(a)$. This definition can also be naturally extended to the case where we consider $\mathcal{S}_\infty$, the set of pmfs over an infinite yet countable alphabet.

To give an intuition of majorization, think of q and p as being non-increasingly sorted pmfs over $[\![1, n]\!]$, i.e $p_j \geq p_{j+1}$ and $q_j \geq q_{j+1}$ for all $j \in [\![1, n-1]\!]$. Then $q \preceq_m p$ means

$$\forall k \in [\![1, n]\!] \ \sum_{i=1}^{k} p_i \geq \sum_{i=1}^{k} q_i \iff \forall k \in [\![1, n]\!] \ F_p(k) \geq F_q(k) \iff F_p \geq F_q$$

where F_p, F_q are the cdfs corresponding to the pmfs p, q respectively. We see that p majorizes q when $F_p \geq F_q$, i.e when F_p grows faster to 1.

In the general, non-sorted case, $q \preceq_m p$ when every $k-$set of largest probabilities of p grows bigger and nearer to 1 than every corresponding $k-$set of largest probabilities of q.

2.2.3 Definition of aggregation

Given 2 pmfs p, q on supports $\mathcal{P}$ and $\mathcal{Q}$ respectively, p is an **aggregation** of q if $\exists g : \mathcal{Q} \to \mathcal{P}$ such that $X \sim q$ implies $g(X) \sim p$. Equivalently, p is an aggregation of q when there exists a partition $\{I_j\}_{j \in \mathcal{P}}$ of $\mathcal{P}$ such that $p_j = \sum_{i \in I_j} q_i \ \forall j \in \mathcal{P}$.
We write $q \sqsubseteq p$.

2.2.4 Aggregation implies majorization

Let $X \sim p_X$ and $Y \sim p_Y$ be discrete RVs. Then

$$p_X \sqsubseteq p_Y \implies p_X \preceq_m p_Y.$$

This is proven in [8]. The idea is that, since each $p_Y(y)$ is the sum of multiple at least 1 probability mass $p_X(x)$ - i.e $p_Y(y_i) = \sum_{j \in \mathcal{I}_i} p_X(x_j)$ with $|\mathcal{I}_i| \geq 1$, the sum of the k biggest probability masses of p_Y equates at least the k biggest probability masses of p_X.

2.2.5 Partial converse

Let $\text{Geom}(x; \gamma) = \gamma(1-\gamma)^{x-1}$ and $\text{Geom}(\gamma)$ be the corresponding pmf. Given two pmfs p over $\mathcal{P}$ and q over $\mathcal{Q}$, we define

$$\forall x \in \mathcal{P} \ \forall y \in \mathcal{Q} \ (p \times q)(x, y) = p(x)q(y).$$

Then

$$p_X \preceq_m p_Y \implies p_X \times \text{Geom}\left(\frac{1}{2}\right) \sqsubseteq p_Y.$$

This result is proven in [4].

2.3 Properties of cdfs

2.3.1 Stochastic dominance

A cdf F_1 **stochastically dominates** another cdf F_2 if $F_1(t) \leq F_2(t) \ \forall t \in \mathbb{R}$, i.e $1 - F_1 \geq 1 - F_2$. We write $F_1 \leq F_2$.
Equivalently, from [9]: a cdf F_1 dominates another cdf F_2 if and only if $\mathbb{E}_{X \sim F_1}(u(X)) \geq \mathbb{E}_{X \sim F_2}(u(X))$ for all u non-decreasing, conditioned on the existence of $\mathbb{E}_{X \sim F_1}(|X|)$ and $\mathbb{E}_{X \sim F_2}(|X|)$.

2.3.2 Convolution of cdfs

Let $\mathcal{P}_+$ be the space of cdfs over $[0; +\infty)$. Letting $F_1, F_2 \in \mathcal{P}_+$, define the **convolution** of cdfs

$$(F_1 * F_2)(t) = \int_0^t F_1(t-s) dF_2(s).$$

Concretely $F_1 * F_2$ is the cdf of $X + Y$ where $X \sim F_1$, $Y \sim F_2$ and $X \ \& \ Y$ are independent.

2.4 Distances

Let F_1, F_2 be 2 cdfs, and let $\mathcal{M}(\mathbb{R})$ be the collection of measurable subsets of $\mathbb{R}$.

2.4.1 Total variation

The **total variation distance** reads:

$$\begin{aligned} d_{\mathrm{TV}}(F_1, F_2) &= \sup_{A \in \mathcal{M}(\mathbb{R})} \left| \mathbb{P}\left(X \in A | X \sim F_1\right) - \mathbb{P}\left(X \in A | X \sim F_2\right) \right| \\ &= \sup_{A \in \mathcal{M}(\mathbb{R})} \left| \int \mathbb{1}\left\{t \in A\right\} dF_1(t) - \int \mathbb{1}\left\{t \in A\right\} dF_2(t) \right|. \end{aligned}$$

2.4.2 Uniform distance

The **uniform metric** reads

$$\begin{aligned} d_U(F_1, F_2) &= \sup_{x \in \mathbb{R}} |F_1(x) - F_2(x)| \\ &= ||F_1 - F_2||_\infty . \end{aligned}$$

3 Information spectrum and information majorization

This section delves deeper into section II of [2]. It introduces both common and new tools specific to this area of research, namely the information spectrum and a generalization of the common definition of information majorization. The latter will involve the so-called G_F function which we will need to understand deeply. The section ends with a statement for approximating arbitrary cdfs by so-called information spectrum cdfs.

3.1 Information spectrum

3.1.1 Definition

Define the **self information** of X as the function

$$\iota_X(x) = -\log(p_X(x)).$$

We denote F_{ι_X} to be the cdf of $\iota_X(X)$, and F_{ι_X} is of the form:

$$F_{\iota_X}(x) = \mathbb{P}\left(\iota_X(X) \leq x\right) = \mathbb{P}\left(-\log(p_X(X)) \leq x\right) = \mathbb{P}\left(\log(p_X(X)) \geq -x\right) = \mathbb{P}\left(p_X(X) \geq 2^{-x}\right).$$

F_{ι_X} is the **information spectrum** of X.
Lastly, we define the **set of information spectrum cdfs** as

$$\mathcal{P}_\iota = \left\{F_{\iota_X} \mid X \text{ is a discrete RV}\right\}.$$

Observe that, when X is a discrete RV,

$$H(X) = \mathbb{E}\left(-\log(p_X(X))\right) = \mathbb{E}\left(\iota_X(X)\right) = \mathbb{E}\left(F_{\iota_X}\right).$$

3.1.2 Examples

- X constant such that $p_X(c) = \mathbb{P}\left(X = c\right) = 1$ and thus $\iota_X(X) = -\log(1) = 0$, and $F_{\iota_X}(x) = \mathbb{1}\left\{x \geq 0\right\}$. Note that $\iota_X(X)$ is constant, but this result does not depend on c.
- $X \sim \mathrm{DisUnif}(\llbracket 1, n \rrbracket)$: $p_X(x) = \frac{1}{n}\mathbb{1}\left\{x \in \llbracket 1, n \rrbracket\right\}$, yielding $p_X(X) = \frac{1}{n}$, and $F_{\iota_X}(x) = \mathbb{1}\left\{x \geq \log(n)\right\}$. $\iota_X(X) = -\log\left(\frac{1}{n}\right) = \log(n)$, which is again constant.
- $X \sim \mathrm{Geom}(p)$, $p_X(x) = (1-p)^{x-1}p$. In the previous cases, $p_X(X)$ was a constant random variable, which is no longer true here - in fact p_X is an injective function. We have

  $$\iota_X(X) = -\log((1-p)^{X-1}p) = -\log(p) - (X-1)\log(1-p) = aX + b$$

  with $a = -\log(1-p)$ and $b = \log(1-p) - \log(p)$. Hence, here, $\iota_X(X)$ is a scaled and shifted geometric random variable.

3.2 Information majorization

Given a cdf F, recall the definition of the quantile function F^{-1}

$$F^{-1}(t) = \inf\{x \geq 0 \mid F(x) \geq t\}.$$

Let $F_1, F_2 \in \mathcal{P}_+$. F_1 is **informationally majorized** by F_2 i.e $F_1 \overset{\iota}{\preceq} F_2$ if

$$G_{F_1}(\gamma) \geq G_{F_2}(\gamma) \ \forall \gamma \in [0;1]$$

where

$$G_F(\gamma) = \int_0^\gamma 2^{F^{-1}(s)} ds = \int_{\mathbb{R}} 2^t d\min\{F(t), \gamma\}.$$

Note that $F^{-1}(\gamma) = \log G'_F(\gamma)$ where G'_F is the left derivative of G_F, and G_F is continuous, strictly increasing, convex and $G_F(\gamma) \geq \gamma$ with $G_F(0) = 0$.

Special case of information spectrum cdfs: consider a random variable $X \sim p_X$ with support $\mathcal{P}$. Then define $g_{p_X}(t)$ over $\mathbb{R}_+$ by interpolating linearly between values of $g_{p_X}(k) = \max_{A \subseteq \mathcal{P}:\ |A| \leq k} p_X(A)$ where $k \in \mathbb{N}$. We will see that

$$G_{F_{\iota_X}}(\gamma) = g_{p_X}^{-1}(\gamma) = \int_0^\gamma 2^{F_{\iota_X}^{-1}(s)} ds.$$

So we see that information majorization is a generalization of the definition of majorization, presented above in 2.2.2.

3.3 Dealing with G_F

Understanding this function is the very core of the contribution of [2]. Hence we take some time to reflect upon its various interpretations.

3.3.1 G_F integral as expectation

Let $F \in \mathcal{P}_+$. Then

$$G_F(\gamma) = \int_{\mathbb{R}} 2^t d\min\{F(t), \gamma\} = \inf_{p_{Q|T}: Q \in [0;1], \mathbb{E}(Q) = \gamma, T \sim F} \mathbb{E}\left(2^T Q\right).$$

This forces the appearance of $2^{T_1+T_2}$ in subsequent computations, which are useful to study convolutions of some distributions. The above holds because

$$G_F(\gamma) = \int_{\mathbb{R}} 2^t d\min\{F(t), \gamma\} = \int_0^\gamma 2^{F^{-1}(s)} ds = \int_0^{F^{-1}(\gamma)} 2^x dF(x) = \mathbb{E}_{T \sim F}\left(2^T \mathbb{1}\left\{0 \leq T \leq F^{-1}(\gamma)\right\}\right).$$

Note that $Q_T = \mathbb{1}\left\{0 \leq T \leq F^{-1}(\gamma)\right\}$ has an expected value $F(F^{-1}(\gamma)) = \gamma$ and is in $\{0;1\} \subseteq [0;1]$. It is the minimizer of $\mathbb{E}\left(2^T Q\right)$ for the class of considered Q's as other Q's set a nonzero weight to the expression $2^T Q$ for "large" T's, larger than $F^{-1}(\gamma)$.

3.3.2 $G_{F_{\iota_X}}$ as a counter

Given a discrete RV X, $G_{F_{\iota_X}}(\gamma)$ counts the linearly interpolated number of entries in p_X so that the total probability mass of the chosen entries is γ. In other words, $G_{F_{\iota_X}}(\gamma) = \int_0^\gamma 2^{F_{\iota_X}^{-1}(s)} ds$ gives the total count per probability up to the $\gamma-$quantile of F_{ι_X}, giving the linearly interpolated count of the largest set of entries of p_X that sum to γ.

Mathematically, by an alternate definition of the inverse of g_{p_X},

$$\begin{aligned} G_{F_{\iota_X}}(\gamma) &= g_{p_X}^{-1}(\gamma) \\ &= \inf\left\{k + \frac{\gamma - g_{p_X}(k)}{g_{p_X}(k+1) - g_{p_X}(k)} \mid k \in \mathbb{N},\ g_{p_X}(k+1) > \gamma\right\} \\ &= k_\gamma + \frac{\gamma - g_{p_X}(k_\gamma)}{g_{p_X}(k_\gamma + 1) - g_{p_X}(k_\gamma)} \end{aligned}$$

where

$$k_\gamma = \max\{k \in \mathbb{N} \mid g_{p_X}(k) \leq \gamma\}.$$

To make things more concrete, assume that $p_X(x+1) \geq p_X(x)$ for all x. Then the formulas above become

$$G_{F_{\iota_X}}(\gamma) = k_\gamma + \frac{\gamma - \sum_{x=1}^{k_\gamma} p_X(x)}{\sum_{x=1}^{k_\gamma+1} p_X(x) - \sum_{x=1}^{k_\gamma} p_X(x)} = k_\gamma + \frac{\gamma - \sum_{x=1}^{k_\gamma} p_X(x)}{p_X(k_\gamma+1)}$$

and

$$k_\gamma = \max\{k \in \mathbb{N} \mid \sum_{x=1}^{k} p_X(x) \leq \gamma\}.$$

3.3.3 Why is $G_{F_{\iota_X}}(\gamma) = g_{p_X}^{-1}(\gamma) = \int_0^\gamma 2^{F_{\iota_X}^{-1}(s)} ds$?

It is by construction that $G_{F_{\iota_X}}(\gamma) = g_{p_X}^{-1}(\gamma)$ and we have understood $G_{F_{\iota_X}}(\gamma)$ as a counter of the probabilities that we need to sum up to reach γ, but why is $g_{p_X}^{-1}(\gamma) = \int_0^\gamma 2^{F_{\iota_X}^{-1}(s)} ds$?
We will prove this in the case that $p_X(x+1) \geq p_X(x)$, and this is w.l.o.g since g_{p_X} only depends on the multiset of probabilities $\{p_X(x) | x \in \mathbb{N}\}$ and not on its ordering.
Pick an integer $k_\gamma \in \mathbb{N}$, and define $F_X(k_\gamma) = \gamma$, equivalently $F_X^{-1}(\gamma) = k_\gamma$. Also assume that $p_X(x)$ is non-decreasing, w.l.o.g. Then

$$\begin{aligned}
\int_0^\gamma 2^{F_{\iota_X}^{-1}(s)} ds &= \sum_{x=1}^{k_\gamma} \int_{F_X(x-1)}^{F_X(x)} 2^{F_{\iota_X}^{-1}(s)} ds \\
&= \sum_{x=1}^{k_\gamma} \int_{F_X(x-1)}^{F_X(x)} 2^{\inf\{t \mid \mathbb{P}(-\log(p_X(X) \leq t) \geq s\}} ds \\
&= \sum_{x=1}^{k_\gamma} \int_{F_X(x-1)}^{F_X(x)} 2^{\inf\{t \mid \mathbb{P}\left(p_X(X) \geq 2^{-t}\right) \geq s\}} ds \\
&= \sum_{x=1}^{k_\gamma} \int_{F_X(x-1)}^{F_X(x)} 2^{\inf\{-\log(l) | \mathbb{P}(p_X(X) \geq l) \geq s\}} ds \\
&= \sum_{x=1}^{k_\gamma} \int_{F_X(x-1)}^{F_X(x)} 2^{\inf\{-\log(l) | \mathbb{P}(p_X(X) \geq l) \geq F_X(x)\}} ds
\end{aligned}$$

The last step follows because $\inf\{-\log(l) | \mathbb{P}(p_X(X) \geq l) \geq s\}$ is constant over $]F_X(x-1), F_X(x)[$ since the distribution of X does not change on that interval. Then, by non-decreasingness of p_X

$$\begin{aligned}
\int_0^\gamma 2^{F_{\iota_X}^{-1}(s)} ds &= \sum_{x=1}^{k_\gamma} \int_{F_X(x-1)}^{F_X(x)} 2^{\inf\{-\log(l) | \mathbb{P}(p_X(X) \geq l) \geq F_X(x)\}} ds \\
&= \sum_{x=1}^{k_\gamma} \int_{F_X(x-1)}^{F_X(x)} 2^{-\log(p_X(x))} ds \\
&= \sum_{x=1}^{k_\gamma} \frac{F_X(x) - F_X(x-1)}{p_X(x)} \\
&= \sum_{x=1}^{k_\gamma} \frac{p_X(x)}{p_X(x)} \\
&= k_\gamma.
\end{aligned}$$

The proof can be easily adapted to the case where it is not necessarily the case that $F_X(k) = \gamma$ for some $k \in \mathbb{N}$, by adding the following integral in the sum:

$$\int_{F_X(k_\gamma)}^{\gamma} 2^{-\log(p_X(k_\gamma+1)} ds = \frac{\gamma - F_X(k_\gamma)}{F_X(k_\gamma+1) - F_X(k_\gamma)} = \frac{\gamma - F_X(k_\gamma)}{p_X(k_\gamma+1)}$$

where $k_\gamma = \max\{k \in \mathbb{N} | F_X(k) \leq \gamma\}$, justifying the interpretation of $G_{F_{\iota_X}}(\gamma) = \int_0^\gamma 2^{F_{\iota_X}^{-1}(s)} ds$ as the smallest linearly interpolated count of probabilities $p_X(x)$ that sum up to γ.

3.4 Properties of information majorization

We quickly list a few basic properties of $\overset{\iota}{\preceq}$: propositions 1, 2 and 3 from [2].
Let $F_1, F_2, F_3 \in \mathcal{P}_+$.

1. $F_1 \leq F_2 \implies F_1 \overset{\iota}{\preceq} F_2$
2. $F_1 \overset{\iota}{\preceq} F_2 \implies \mathbb{E}(F_1) \geq \mathbb{E}(F_2)$
3. $F_1 \overset{\iota}{\preceq} F_2 \wedge F_3 \overset{\iota}{\preceq} F_4 \implies F_1 * F_3 \overset{\iota}{\preceq} F_2 * F_4 \wedge (1-\lambda)F_1 + \lambda F_3 \overset{\iota}{\preceq} (1-\lambda)F_3 + \lambda F_4 \ (\lambda \in [0;1])$

For two discrete RVs X, Y, we have:

$$p_X \preceq_m p_y \iff F_{\iota_X} \overset{\iota}{\preceq} F_{\iota_Y}$$

The core steps of the proofs of these properties come down to understanding G_F and its interpretations given above.

3.5 Approximation by an information spectrum cdf

Section II of [2] leads up to the present proposition, which essentially says that every discrete cdf (in $\mathcal{P}_+$) is close, in a sense, to an information spectrum cdf. Indeed, the expectation of any cdf F over the positive reals can be approximated by the expectation of an information spectrum cdf F_{ι_X}, which is also the entropy of X, $H(X)$. Furthermore, X and F are related through $F_{\iota_X} \overset{\iota}{\preceq} F$.

3.5.1 Proposition 4 in [2]

Let $F \in \mathcal{P}_+$. Then

$$\exists X \text{ discrete RV such that } F_{\iota_X} \overset{\iota}{\preceq} F \text{ and } \mathbb{E}(F) \leq H(X) = \mathbb{E}(F_{\iota_X}) \leq \mathbb{E}(F) + \underbrace{h_2\left(\min\left\{\frac{1}{2}, \sqrt{\frac{\mathbb{E}(F)}{\log(e)}}\right\}\right)}_{\leq 1}$$

3.5.2 Ideas of the proof

- We want to construct X such that $\mathbb{E}(F) \leq H(X)$, and then find an upper bound to $H(X)$. Thus, from our previous findings in 3.4, we seek X such that $F_{\iota_X} \overset{\iota}{\preceq} F$.
- Equivalently, we want $G_{F_{\iota_X}} \geq G_F$, and $G^{-1}_{F_{\iota_X}}$ needs to be piecewise affine, in this case affine in each interval $[G^{-1}_{F_{\iota_X}}(k-1), G^{-1}_{F_{\iota_X}}(k)]$. Thus we linearly interpolate $(G_F^{-1}(k), k)$ for $k \in \mathbb{N}$.
- We are interested in $q_k = G_F^{-1}(k) - G_F^{-1}(k-1)$, which can be shown to be non-increasing by concavity of G_F^{-1}, and set these weights to be those of the pmf of a random variable X: $p_X(k) = q_k, k \geq 1$. $G_{F_{\iota_X}}$ will have the desired properties mentioned in the previous point.
- The upper-bound is the consequence of some usual calculus steps as well as $F^{-1}(\gamma) = \log G'_F(\gamma)$ and the log sum inequality $\sum_{i=1}^n a_i \log\left(\frac{a_i}{b_i}\right) \geq (\sum_{i=1}^n a_i) \log\left(\frac{\sum_{i=1}^n a_i}{\sum_{i=1}^n b_i}\right)$. The latter is derived from the Jensen inequality, details in [10].

3.5.3 Example 1 - Constant random variable

Let us try to find a suitable discrete random variable X for some cdfs $F \in \mathcal{P}_+$, i.e X discrete such that $\mathbb{E}(F) \leq H(X) \leq \mathbb{E}(F) + 1$ and $F_{\iota_X} \overset{\iota}{\preceq} F$.
To start with a trivial example: let $F(t) = \mathbb{1}\{t \geq 1\}$ be the cdf of the constant random variable Y such that $\mathbb{P}(Y = 1) = 1$. Then trivially, with $X \sim \text{Ber}\left(\frac{1}{2}\right)$ which is indeed a discrete distribution, we indeed have that $H(X) = \log_2(2) = 1$ and thus $1 = \mathbb{E}(F) \leq H(X) \leq \mathbb{E}(F) + 1$. We check that $F_{\iota_X} \overset{\iota}{\preceq} F$:

$$F_{\iota_X}(t) = \mathbb{P}(-\log(p_X(X)) \leq t) = \mathbb{1}\{t \geq \log(2)\} = \mathbb{1}\{t \geq 1\}.$$

So we have $F_{\iota_X} = F$, thus $F_{\iota_X} \overset{\iota}{\preceq} F$.

3.5.4 Example 2 - Exponential random variable

Perhaps more interestingly, let $Y \sim \exp(\ln(2))$. a suitable X is not as obvious, so we construct it by following the steps of the proof. Note that $\mathbb{E}(Y) = \ln(2)$.
First we need to find G_F, using $\log(G_F') = F^{-1}$. Here, $F_Y(t) = F(t) = 1 - e^{-\ln(2)t} = 1 - 2^{-t}$, and

$$\forall \gamma \in [0,1[\ 1-2^{-t} = \gamma \iff t = \log_2\left(\frac{1}{1-\gamma}\right) \implies F^{-1}(\gamma) = \log\left(\frac{1}{1-\gamma}\right) \implies G_F'(\gamma) = \frac{1}{1-\gamma}$$

and thus

$$G_F(\gamma) = \int_0^\gamma \frac{1}{1-s} ds = \int_{1-\gamma}^1 \frac{1}{u} du = \ln\left(\frac{1}{1-\gamma}\right).$$

To construct the pmf of X, we need to find $\gamma_k = G_F^{-1}(k)$ for all $k \in \mathbb{N}$

$$\ln\left(\frac{1}{1-\gamma_k}\right) = k \iff \frac{1}{1-\gamma_k} = e^k \iff 1-\gamma_k = e^{-k} \iff \gamma_k = 1 - e^{-k}.$$

Also set $\gamma_{-1} = -1$ as in the proof. Then define

$$\forall k \in \mathbb{N}^*\ p_k = \gamma_k - \gamma_{k-1} = e^{-k+1} - e^{-k} = e^{-(k-1)}\frac{e-1}{e} = e^{-(k-1)}(1-e^{-1}),$$

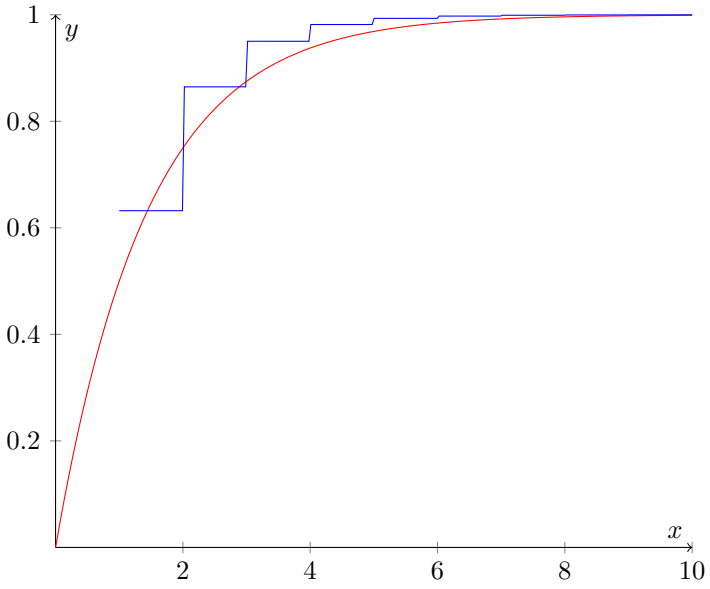
hence $X \sim \text{Geom}(1-e^{-1})$, and:

$$H(X) = \frac{-e^{-1}\ln(e^{-1}) - (1-e^{-1})\ln(1-e^{-1})}{1-e^{-1}} = \frac{h_2(1-e^{-1})}{1-e^{-1}}.$$

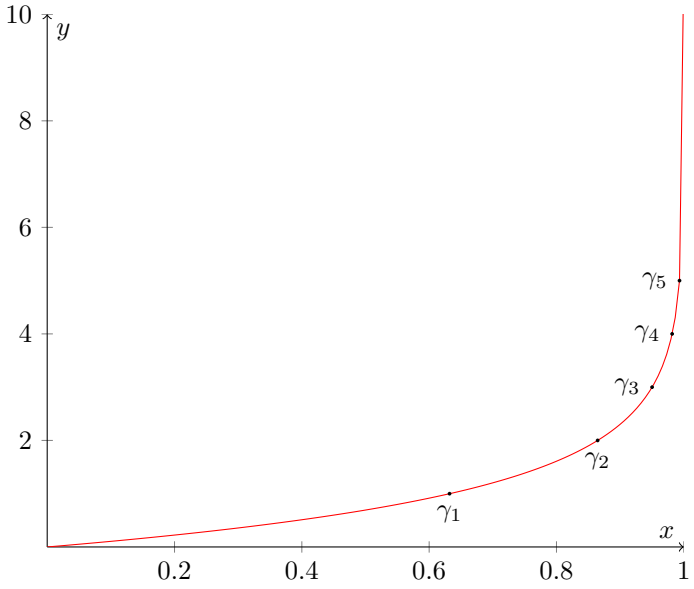
Lastly, we check that $\mathbb{E}(Y) \leq H(X) \leq \mathbb{E}(Y) + 1$

$$\mathbb{E}(Y) = \ln(2) \approx 0.69 \leq H(X) = \frac{h_2(1-e^{-1})}{1-e^{-1}} \approx 1.04 \leq \mathbb{E}(Y) + 1 = 1 + \ln(2) \approx 1.69$$

Plotting $F(t) = (1-2^{-t})\mathbb{1}\{t \geq 0\}$ in red against $F_X(t) = (1-e^{-\lfloor t \rfloor})\mathbb{1}\{t \geq 1\}$ in blue:



Plotting $G_F(\gamma) = \ln\left(\frac{1}{1-\gamma}\right)$ along with some points $(G_F^{-1}(k), k) = (\gamma_k, k)$ at which G_F is interpolated to construct $G_{F_{\iota_X}}$:



4 Kolmogorov's infinite divisibility

Kolmogorov studies the concept of divisibility of randomness in his 1956 paper *Two uniform limit theorems for sums of independent random variables* [3], which we recap here. This is the main motivation behind the present work. Beyond this section, appendix B gives some more examples and cool properties of Kolmogorov infinite divisibility.

4.1 Definition of infinite divisibility

The definition from [3] is more general than in [2], but we use the definition from [2] as it is the most common nowadays: the cdf F is infinitely divisible if $\forall n \in \mathbb{N}\ \exists F_n$ such that, when $X \sim F$, then $\exists \{Z_i\}_{i=1}^n \overset{\text{i.i.d}}{\sim} F_n$ such that $X \sim \sum_{i=1}^n Z_i$. In other words, $F = F_n^{*n} = F_n * F_n * \ldots * F_n$ n times. We say that F is divisible n times, or $n-$divisible.

The way to understand this definition is to picture the randomness in X as having been diluted into n parts Z_i which are created equal (identical distribution) and independently. Observe that, since the resulting RV X has a cdf F which is an $n-$convolution for all n, F is required to be quite smooth.

4.2 Kolmogorov uniform limit theorem

The second uniform limit theorem from [3] states: $\exists C > 0$ such that $\forall n \in \mathbb{N}\ \forall F$ cdf divisible n times $\exists \tilde{F}$ infinitely divisible (as defined above) such that

$$\left|\left|F - \tilde{F}\right|\right|_\infty \leq C n^{-\frac{1}{5}}.$$

This bound has been improved since Kolmogorov's statement to $Cn^{-\frac{2}{3}}$ [11].

4.3 Examples

4.3.1 Compound Poisson random variables are infinitely divisible

Let $V = \beta + \sum_{i=1}^N Z_i$ where $N \sim \text{Poi}(\alpha)$ and $\{Z_i\} \overset{\text{i.i.d}}{\sim} Z$, $\beta \in \mathbb{R}$. Then V is infinitely divisible, essentially thanks to the infinite divisibility of the Poisson distribution.
Indeed, let $n \in \mathbb{N}$, then letting $N = N_1 + \ldots + N_n$ with $\{N_j\}_{j=1}^n \overset{\text{i.i.d}}{\sim} \text{Poi}\left(\frac{\alpha}{n}\right)$ and relabelling Z_i as $Z_i^k = Z_{i+\sum_{j=1}^{k-1} N_j}$,

$$V = \left(\frac{\beta}{n} + \ldots + \frac{\beta}{n}\right) + \sum_{k=1}^n \sum_{i=1}^{N_k} Z_i = \sum_{k=1}^n \left(\frac{\beta}{n} + \sum_{i=1}^{N_k} Z_i^k\right)$$

which divides V into n i.i.d parts $\frac{\beta}{n} + \sum_{i=1}^{N_k} Z_i^k$. We will use this fact later.

4.3.2 More examples

See [12] for full details, e.g necessary/sufficient conditions for infinite divisibility. Distributions with bounded supports are *not* infinitely divisible. Common distributions are very often infinitely divisible: Normal, Gamma, Cauchy, Pareto, Poisson, Geometric, Negative-binomial, etc.

5 Informational infinite divisibility

Here we present section III of [2], where the author formally introduces informational divisibility and states the main results of the paper. The proofs' main complexities shift from calculus to probability concepts.

5.1 Spaces of divisibility, definitions

5.1.1 Classic sense

We define **the space of n-divisible** cdfs over $\mathcal{P}_+$ as $\mathcal{P}_+^{*n}$ and **the space of infinitely divisible** cdfs over $\mathcal{P}_+$ as $\mathcal{P}_+^{*\infty}$ as

$$\mathcal{P}_+^{*n} = \{F^{*n} \mid F \in \mathcal{P}_+\},\ \mathcal{P}_+^{*\infty} = \bigcap_{n\in\mathbb{N}} \mathcal{P}_+^{*n}.$$

5.1.2 Information sense

Recall $\mathcal{P}_\iota = \{F_{\iota_X} \mid X \text{ is a discrete RV}\}$.
We similarly define **the space of n-informationally divisble** cdfs $\mathcal{P}_\iota^{*n}$ and **the space of informationally infinitely divisible** cdfs $\mathcal{P}_\iota^{*\infty}$ as

$$\mathcal{P}_\iota^{*n} = \{F^{*n} \mid F \in \mathcal{P}_\iota\},\ \mathcal{P}_\iota^{*\infty} = \bigcap_{n\in\mathbb{N}} \mathcal{P}_\iota^{*n}.$$

5.1.3 Motivation for the definitions of divisibility in the information sense

In the classic sense, if $X \sim F^{*n} \in \mathcal{P}_+^{*n}$, then $X \sim \sum_{i=1}^n Z_i$ where $\{Z_i\}_{i=1}^n \overset{\text{i.i.d}}{\sim} F$, as in the *randomness* in X is distilled into n identical and independent parts of randomness.

In the information sense, X is replaced by its self-information $\iota_X(X)$, thus we want $\iota_X(X) \sim \sum_{i=1}^n \iota_{Z_i}(Z_i)$, i.e the distribution F_{ι_X} of $\iota_X(X)$ is divisible in the classic sense *by an information spectrum cdf* F_{ι_Z}. Because of this very close tie between probabilistic infinite divisibility and information infinite divisibility, many results from the former are mirrored in the latter.

Equivalently, we can define the $n-$informational divisibility of X by the existence of $\{Z_i\}_{i=1}^n \overset{\text{i.i.d}}{\sim} F$ such that $H(X|Z^n) = H(Z^n|X) = 0$. Thus X carries the same information as Z^n.

5.2 No informationally infinitely divisible discrete random variable

We go over a somewhat surprising statement:

5.2.1 Proposition 5 in [2]

There is no informationally infinitely divisible discrete random variable.

$$\mathcal{P}_\iota^{*\infty} = \emptyset.$$

5.2.2 Lemma - second largest entries

Let Z be a discrete random variable on $\mathbb{N}^*$, and let $\{Z_i\}_{i=1}^n \sim p_Z$. Then either Z is uniform or the number of second largest entries of p_{Z^n} is a multiple of n, i.e the number of vectors z^n which result in the second largest value that $p_{Z^n}(z^n)$ can attain is a multiple of n.
The proof considers the case w.l.o.g of $p_Z(1) \geq p_Z(2) \geq p_Z(3) \geq \ldots$. In the non-uniform case, there exists $a \in \mathbb{N}$ such that $p_Z(1) = p_Z(a) > p_Z(a+1) > 0$. The entries z^n that maximize $p_{Z^n}(z^n)$ are those such that $p_{Z^n}(z^n) = (p_Z(1))^n$. The second largest value of $p_{Z^n}(z^n) = (p_Z(1))^{n-1}p_Z(a+1)$ is attained for entries z^n that contain exactly one z_i such that $p_Z(z_i) = p_Z(a+1)$ and the remaining z_j have $p_Z(z_j) = p_Z(1)$. There is a multiple of n of such entries.

5.2.3 Proof of the main statement

Assume there exists a discrete random variable X such that $F_{\iota_X} \in \mathcal{P}_\iota^{*\infty}$.

- X cannot be uniform: let n be the size of the support of X. $\iota_X(X)$ cannot be divided into $m > n$ i.i.d $\iota_Z(Z_i) \overset{\text{i.i.d}}{\sim} p$.
- Consider Z^n such that $F_{\iota_X} = F_{\iota_{Z^n}}$, which exists since $F_{\iota_X} \in \mathcal{P}_\iota^{*\infty} \subset \mathcal{P}_\iota^{*n}$, with n larger than the number of second largest entries in p_X. But since X is not uniform, Z_i cannot be uniform, since uniform Z_i would lead to uniform X. Thus the number of second largest entries of p_{Z^n}, which is the same as for p_X is a multiple of n, which leads to a contradiction.
- By the previous lemma, these are the only 2 cases we have to consider.

5.2.4 So what now?

The previous proposition may leave us perplexed. It is the answer, albeit dissatisfying, to the initial question the author had about the infinite divisibility of information: no, (discrete) information is not infinitely divisible.

However, (discrete) information may be at least approximately $n-$informationally divisible, i.e we might be able to approximately divide cdfs in $\mathcal{P}_\iota$ into $n-$information spectrum cdfs, which amounts to essentially projecting a cdf in $\mathcal{P}_\iota$ into $\mathcal{P}_\iota^{*n}$. Effectively, this would solve the initial engineering problem we had about secret distributed storage.

The author proceeds in two steps, by first bounding the gap between $\mathcal{P}_\iota$ and $\mathcal{P}_+^{*n}$, and then the gap between $\mathcal{P}_\iota$ and $\mathcal{P}_\iota^{*n}$, noticing that $\mathcal{P}_\iota^{*n} \subset \mathcal{P}_+^{*n}$. These are the next two subsections.

Before proceeding, it may be of use to specify the case of continuous random variables. Indeed, $\mathcal{P}_\iota$ only contains information spectrum cdfs of discrete RVs.
In the continuous case, it suffices to look at $X \sim \text{Unif}([0,1])$. Express X in its binary representation $X = 0.B_1B_2B_3B_4\ldots$, we see that $\{B_i\}_{i=1}^\infty \overset{\text{i.i.d}}{\sim} \text{Ber}\left(\frac{1}{2}\right)$. To informationally divide X into n i.i.d parts $\{Z_i\}_{i=1}^n$, pick $Z_i = 0.B_iB_{i+n}B_{i+2n}\ldots$. Then to informationally divide $X' \sim F$ with F some continuous cdf, it suffices to divide $F(X') \sim \text{Unif}([0,1])$ which contains the same information as X' by bijectivity of F.

5.3 Multiplicative gap between $\mathcal{P}_\iota$ and $\mathcal{P}_+^{*n}$

In fact we will even boun the more general gap between $\mathcal{P}_+$ and $\mathcal{P}_+^{*n}$ for all $n \in \mathbb{N} \cup \{\infty\}$ - notice that $\mathcal{P}_\iota \subset \mathcal{P}_+$. A corollary of the below statement is that we can approximately divide any discrete random variable into equal parts in the classic sense.

5.3.1 Theorem 1 in [2]

For any $F \in \mathcal{P}_+$ and $n \in \mathbb{N} \cup \{\infty\}$, $\exists \tilde{F} \in \mathcal{P}_+^{*n}$ such that $\tilde{F} \leq F$ and

$$\mathbb{E}(F) \leq \mathbb{E}\left(\tilde{F}\right) \leq \frac{1}{1-(1-n^{-1})^n}\mathbb{E}(F).$$

When $n = \infty$ we define $\frac{1}{1-(1-n^{-1})^n} = \frac{e}{e-1}$.

5.3.2 Ideas of the proof

- Consider $\tilde{F}$ to be the cdf of a compound Poisson distribution: $V \sim \tilde{F}$ where $V = \beta + \sum_{i=1}^{N} Z_i$ where $N \sim \text{Poi}(\alpha)$ for $n = \infty$ else $N \sim \text{Bin}(n, p)$ and $Z_i \overset{\text{i.i.d}}{\sim} Z$ with β well chosen. Such $\tilde{F}$ is guaranteed to be infinitely divisible for $n = \infty$, by infinite divisibility of the Poisson distribution. See 4.3.1 for a formal proof.
- Set $Z = F^{-1}(U) - F^{-1}(\zeta)$ and $\beta = F^{-1}(\zeta)$ where $U \sim \text{Unif}(\zeta, 1)$, this is motivated by the wrong but almost correct choice of setting $Z \sim F$ and $\beta = F^{-1}(1)$ correcting for the case where $N = 0$, since a Poisson random variable contains 0 in its support.
- Set $N \sim \text{Poi}(-\ln(\zeta))$ for $n = \infty$, and $N \sim \text{Bin}(n, 1 - \zeta^{\frac{1}{n}})$ else, independent of Z_i. This yields $F_V \in F_+^{*n}$.
- Define $V' = \mathbb{1}\{N = 0\} F^{-1}(\tilde{U}) + \mathbb{1}\{N \geq 1\} F^{-1}(U)$ where $\tilde{U} \sim \text{Unif}(0, \zeta)$. It can be shown that $V' \sim F$, and it is easy to see that $V \geq V'$, thus $\tilde{F} \leq F$.
- From $\tilde{F} \leq F$ we conclude $\mathbb{E}(F) \leq \mathbb{E}\left(\tilde{F}\right)$. The upper bound is a consequence of easy manipulations of the expectation, noticing that Z is essentially a function of a uniform random variable, thus simplifying the computation of $\mathbb{E}(Z)$.
- Lastly we pick $\zeta = (1 - n^{-1})^n$ for the finite case and $\zeta = e^{-1}$ when $n = \infty$.

5.3.3 Unif(0,1) variable example

Let $W \sim \text{Unif}(0, 1)$, thus $F(t) = F_W(t) = t$ for $t \in [0, 1]$. $F \notin \mathcal{P}_+^{*\infty}$.
We look for a suitable $\tilde{F}$ with $n = \infty$, following the proof. Here, note that $F^{-1}(\gamma) = \gamma$ and $\mathbb{E}(F) = \frac{1}{2}$. Hence, $Z = F^{-1}(U) - F^{-1}(e^{-1}) = U - e^{-1}$, $\beta = e^{-1}$ and $N \sim \text{Poi}(1)$. We get, with $\{U_i\} \overset{\text{i.i.d}}{\sim} \text{Unif}(e^{-1}, 1)$

$$V = e^{-1} + \sum_{i=1}^{N} \left(U_i - e^{-1}\right) = (1 - N)e^{-1} + \sum_{i=1}^{N} U_i.$$

Thus, with $\mathbb{E}(N) = 1$

$$\mathbb{E}(V) = \mathbb{E}(U) = \frac{e^{-1} + 1}{2}$$

and indeed $\mathbb{E}(V) = \frac{e^{-1}+1}{2} \approx 0.683 \geq \frac{1}{2}$, then for the upper bound

$$\frac{e}{e-1}\frac{1}{2} = \frac{e}{2e-2} \approx 0.791 \geq \mathbb{E}(V).$$

5.3.4 DisUnif$(1, m)$ variable example

Let $W \sim \text{DisUnif}(1, m)$.
Here $F_W(t) = F(t) = \sum_{i=1}^{m} \frac{1}{m} \mathbb{1}\{t \geq i\}$. Thus $F^{-1}(\frac{i}{m}) = i$ for $i \in [\![1, m]\!]$. It follows from the definition of F^{-1} that, if $0 < \varepsilon < 1$, then $F^{-1}\left(\frac{i+\varepsilon}{m}\right) = i + 1$ for $i \in [\![0, m-1]\!]$. Thus we conclude $F^{-1}(\gamma) = \lceil m\gamma \rceil$ for $\gamma \in [0, 1]$. Lastly note $\mathbb{E}(F) = \frac{m+1}{2}$.
Let $n = m + k$, $k > 0$ but $k < \infty$. Then with $N \sim \text{Bin}(n, n^{-1})$, $Z = F^{-1}(U) - F^{-1}((1 - n^{-1})^n) = \lceil mU \rceil - \lceil m(1 - n^{-1})^n \rceil$, with $U \sim \text{Unif}((1 - n^{-1})^n, 1)$ we have

$$V = (1 - N)\lceil m(1 - n^{-1})^n \rceil + \sum_{i=1}^{N} \lceil mU_i \rceil.$$

Thus, still with $\mathbb{E}(N) = 1$,

$$\begin{aligned}
\mathbb{E}(V) &= \mathbb{E}(\lceil mU_1 \rceil) \\
&= \frac{1}{1-(1-n^{-1})^n} \int_{(1-n^{-1})^n}^{1} \lceil mu \rceil du \\
&= \frac{1}{m(1-(1-n^{-1})^n)} \int_{m(1-n^{-1})^n}^{m} \lceil x \rceil dx \\
&= \frac{1}{m(1-(1-n^{-1})^n)} \left(\int_0^m \lceil x \rceil dx - \int_0^{m(1-(1-n^{-1})^n)} \lceil x \rceil dx \right) \\
&= \frac{1}{m(1-(1-n^{-1})^n)} \left(\frac{m(m+1)}{2} - \frac{\lfloor m(1-(1-n^{-1})^n) \rfloor (\lfloor m(1-(1-n^{-1})^n) \rfloor + 1)}{2} \right) \\
&\quad - \frac{1}{m(1-(1-n^{-1})^n)} \left((m(1-(1-n^{-1})^n) - \lfloor m(1-(1-n^{-1})^n) \rfloor) \lceil m(1-(1-n^{-1})^n) \rceil \right) \\
&= \frac{1}{(1-(1-n^{-1})^n)} \left(\frac{m+1}{2} - \frac{\lfloor m(1-(1-n^{-1})^n) \rfloor (\lfloor m(1-(1-n^{-1})^n) \rfloor + 1)}{2m} \right) \\
&\quad - \frac{1}{m(1-(1-n^{-1})^n)} \left[(m(1-(1-n^{-1})^n) - \lfloor m(1-(1-n^{-1})^n) \rfloor) \lceil m(1-(1-n^{-1})^n) \rceil \right].
\end{aligned}$$

And here again we can numerically check that the expectation lies in the correct interval.

5.4 Bounding the gap between $\mathcal{P}_\iota$ and $\mathcal{P}_\iota^{*n}$

This is the main statement about the $n-$informational divisibility of discrete random variables.

5.4.1 Theorem 2 in [2]

Let X be a discrete random variable and $n \in \mathbb{N}^*$. Then there exist $\{Z_i\}_{i=1}^n \overset{\text{i.i.d}}{\sim} Z$ and a function f such that $X \sim f(Z^n)$ and

$$H(Z) - \frac{1}{1-(1-n^{-1})^n} \frac{H(X)}{n} \leq \min \left\{ 2.43, h_2 \left(\min \left\{ \frac{1}{2}, \sqrt{\frac{e}{(e-1)\log(e)} \frac{H(X)}{n}} \right\} \right) + h_2(2^{-\frac{1}{n}}) + 2(1-2^{-\frac{1}{n}}) \right\}$$

Essentially,

$$H(Z) = \frac{e}{e-1} \frac{H(X)}{n} + O\left(\frac{\log(n)}{\sqrt{n}} \right).$$

Note that the condition $X \sim f(Z^n)$ is equivalent to $H(X|Z^n) = 0$ since we can always set $X = f(Z^n)$.

5.4.2 Ideas of the proof

- Applying 5.3, there exists $\tilde{F} \in \mathcal{P}_+^{*n}$ such that $\tilde{F} \leq F_{\iota_X}$, hence $\tilde{F} \overset{\iota}{\preceq} F_{\iota_X}$ (3.4). Also $\mathbb{E}\left(\tilde{F}\right) \leq \frac{1}{1-(1-n^{-1})^n} \mathbb{E}(F_{\iota_X}) = \frac{1}{1-(1-n^{-1})^n} H(X)$.
- Let $F \in \mathcal{P}_+$ s.t $F^{*n} = \tilde{F}$. By 3.5.1 there exists a random variable Y such that $F_{\iota_Y} \overset{\iota}{\preceq} F$ and $H(Y) \leq \mathbb{E}(F) + h_2\left(\min\left\{\frac{1}{2}, \sqrt{\frac{\mathbb{E}(F)}{\log(e)}}\right\}\right)$.
- Also, by 3.4 we have that $F_{\iota_Y}^{*n} \overset{\iota}{\preceq} F^{*n} = \tilde{F} \overset{\iota}{\preceq} F_{\iota_X}$ and thus $p_Y^{\times n} \preceq_m p_X$, which implies by 2.2.5 that $p_Y^{\times n} \times \text{Geom}\left(\frac{1}{2}\right) \sqsubseteq p_X$.
- Define $p_B(k) = (1-2^{-k})^{\frac{1}{n}} - (1-2^{-(k-1)})^{\frac{1}{n}}$. Hence we can show that $p_B(k) \sqsubseteq \text{Geom}\left(\frac{1}{2}\right)$. Letting $p_Z = p_Y \times p_B$, we have $p_Z^{\times n} \sqsubseteq p_X$.
- Now what is left to do is to bound $H(Z) = H(Y) + H(B) \leq \mathbb{E}(F) + h_2\left(\min\left\{\frac{1}{2}, \sqrt{\frac{\mathbb{E}(F)}{\log(e)}}\right\}\right) + H(B)$, by using the previous observations.

5.4.3 Discussion

The result proven here is, in some ways, disappointing. Indeed, with $X = f(Z^n)$ where Z_i are i.i.d, the above statement implies

$$H(Z^n) = \sum_{i=1}^{n} H(Z_i) = \frac{e}{e-1} H(X) + O(\sqrt{n}\log(n)).$$

From an engineering perspective, this means that if we want to divide $H(X)$ bits of information into n i.i.d parts, it costs $\frac{e}{e-1}H(X) + O(\sqrt{n}\log(n))$ bits of storage in total, i.e an extra $\left(\frac{e}{e-1} - 1\right) H(X) + O(\sqrt{n}\log(n))$ bits.

However, it is also good news in a sense, since the *per-node* additive overhead of storage space $O\left(\frac{\log(n)}{\sqrt{n}}\right)$ goes to 0 and the factor $\frac{e}{e-1} \approx 1.58$ is not too big.

6 Closing the multiplicative gap

Can we make the previous bound better? The answer is yes, sometimes the factor $\frac{e}{e-1}$ can be improved for free, as we will see in 6.2. Also, it can be improved by imposing more structure on X. This is explained in 6.5.

6.1 Spectral infinite divisibility

A discrete rv X is **spectral infinitely divisible (SID)** if $F_{\iota_X} \in \mathcal{P}_+^{*\infty}$, i.e if its information spectrum is infinitely divisible *in the Kolmogorov sense*.

This is a compromise. We still want to divide $\iota_X(X)$ into n pieces, but these pieces can follow any distribution and not necessarily that of an information spectrum. So we can only write $\iota_X(X) = \sum_{i=1}^n W_i$ instead of $\iota_X(X) = \sum_{i=1}^n \iota_{Z_i}(Z_i)$.
Distributions which are SID: discrete uniform, geometric distribution, etc. The author introduces a general distribution which is SID, it is detailed in appendix C.

6.2 Definition - ratio to infinite divisibility

The multiplicative gap $\alpha = \frac{e}{e-1}$ computed earlier in 5.3 is an upper bound. Sometimes, a better multiplicative gap α can be found, and the following definition formalizes the notion of best multiplicative gap.

Let $F \in \mathcal{P}_+$, its **ratio to infinite divisibility** is

$$r_{\mathrm{ID}}(F) = \begin{cases} \frac{1}{\mathbb{E}(F)} \inf\{\mathbb{E}\left(\tilde{F}\right) \mid \tilde{F} \in \mathcal{P}_+^{*\infty}, \tilde{F} \leq F\} & \text{if } \mathbb{E}(F) \neq 0, \\ 1 & \text{if } \mathbb{E}(F) = 0. \end{cases}$$

In other words, $r_{\mathrm{ID}}(F)$ finds the "best" distribution $\tilde{F}$ meeting the conditions of 5.3, best in the sense that $\mathbb{E}\left(\tilde{F}\right)$ is as close as possible to $\mathbb{E}(F)$.

6.3 Properties of $r_{\mathrm{ID}}(F)$

Listing properties of r_{ID} (proposition 6 in [2]):

- Bound: for any $F \in \mathcal{P}_+$, $1 \leq r_{\mathrm{ID}}(F) \leq \frac{e}{e-1}$.
- Relation to SID: a random variable X is SID $\iff r_{\mathrm{ID}}(F_{\iota_X}) = 1$. In other words, $F_{\iota_X} \in \mathcal{P}_+^{*\infty} \iff r_{\mathrm{ID}}(F_{\iota_X}) = 1$.
- Convolution: Let $F_1, F_2 \in \mathcal{P}_+$ with positive mean, then

$$r_{\mathrm{ID}}(F_1 * F_2) \leq \frac{r_{\mathrm{ID}}(F_1)\,\mathbb{E}(F_1) + r_{\mathrm{ID}}(F_2)\,\mathbb{E}(F_2)}{\mathbb{E}(F_1) + \mathbb{E}(F_2)}$$

We give some more details of the proof of the last property. The convolution property results from the following fact: let $\tilde{F}_1 \leq F_1$ and $\tilde{F}_2 \leq F_2$ such that $\tilde{F}_1, \tilde{F}_2 \in \mathcal{P}_+^{*\infty}$. Then

$$\tilde{F}_1 * \tilde{F}_2 \leq F_1 * F_2.$$

Thus we have

$$\begin{aligned} r_{\mathrm{ID}}(F_1 * F_2) &\leq \frac{\mathbb{E}\left(\tilde{F}_1 * \tilde{F}_2\right)}{\mathbb{E}(F_1 * F_2)} \\ &= \frac{\mathbb{E}\left(\tilde{F}_1\right) + \mathbb{E}\left(\tilde{F}_2\right)}{\mathbb{E}(F_1) + \mathbb{E}(F_2)}. \end{aligned}$$

Picking $\tilde{F}_1$ and $\tilde{F}_2$ achieving $\frac{\mathbb{E}\left(\tilde{F}_i\right)}{\mathbb{E}(F_i)}$ arbitrarily close to $r_{\mathrm{ID}}(F_i)$ yields the desired bound.

6.4 Refined bound on the gap between $\mathcal{P}_\iota$ and $\mathcal{P}_\iota^{*n}$

With this new definition, we can optimize 5.4 for free as done in proposition 7 of [2]:
Let X be a discrete random variable and $n \in \mathbb{N}^*$. Then there exist $\{Z_i\}_{i=1}^n \overset{\text{i.i.d}}{\sim} Z$ and a function f such that $X \sim f(Z^n)$ and

$$H(Z) - r_{\mathrm{ID}}(F_{\iota_X}) \frac{H(X)}{n} \leq \min\left\{2.43, h_2\left(\min\left\{\frac{1}{2}, \sqrt{\frac{e}{(e-1)\log(e)}\frac{H(X)}{n}}\right\}\right) + h_2(2^{-\frac{1}{n}}) + 2(1 - 2^{-\frac{1}{n}})\right\}.$$

Essentially,

$$H(Z) = r_{\mathrm{ID}}(F_{\iota_X}) \frac{H(X)}{n} + O\left(\frac{\log(n)}{\sqrt{n}}\right).$$

6.5 Case of iid sequences

We look at the case where $X = (Y_1, \ldots, Y_m)$ where $Y_i \overset{\text{i.i.d}}{\sim} p_Y$ as an analog to Kolmogorov's uniform theorem 4.2 that looks at $X = \sum_{i=1}^m Y_i$ and the convergence rate of its cdf to an infinitely divisible one. The author proves $r_{\mathrm{ID}}(F_{\iota_{Y^m}}) \to 1$ uniformly. This together with the refined gap on informational infinite divisibility 6.4 yields that, asymptotically, X can be divided into n pieces, with $n \geq m$, with entropy close to the lower bound.

6.6 Theorem 3 in [2]

Let $F \in \mathcal{P}_+^{*m}$, $m \geq 2$. Then there exists $\tilde{F} \in \mathcal{P}_+^{*\infty}$ such that $\tilde{F} \leq F$ and

$$\mathbb{E}\left(\tilde{F}\right) \leq \left(1 + 4.71\sqrt{\frac{\log(m)}{m}}\right) \mathbb{E}(F).$$

Said otherwise,

$$\sup_{F \in \mathcal{P}_+^{*m}} r_{\mathrm{ID}}(F) \leq 1 + 4.71\sqrt{\frac{\log(m)}{m}}.$$

6.7 Conjecture

The author conjectures that, by optimizing a step of the proof in the above theorem, one may potentially prove that there exists $c \in \mathbb{R}$ such that for all $m \in \mathbb{N}$

$$\sup_{F \in \mathcal{P}_+^{*m}} r_{\mathrm{ID}}(F) \leq 1 + 4.71\frac{c}{\sqrt{m}}.$$

This is due to the factor $1 + \frac{2\gamma}{m}$ in lemma 1 of [2] which might potentially be unnecessary, according to the author. We were not able to prove nor disprove his conjecture.

7 Conclusion

As it turns out, information is indeed not infinitely divisible - at least not in the sense defined by [2] - but it is approximately the case, up to some information redundancy factor $\leq \frac{e}{e-1}$. In some cases, notably inspired by [3], information is even asymptotically infinitely divisible.

An interesting modification to the core assumptions made in the paper would be to remove the hypothesis that the Z_i's dividing X need to have the same distribution. Thus a future axis of research in the same vein may be the following problem:

$$\inf_{p_{Z_1} p_{Z_2} \cdots p_{Z_n}} \max_{i=1}^{n} \left| \frac{H(X)}{n} - H(Z_i) \right|$$

where $X = f(Z^n)$ and the Z_i are independent, rather than i.i.d.

Beside the main results, the paper gives important tools and concepts along with fundamental ideas of their interactions. These may prove useful for further research in information and probability theory.

8 Acknowledgements

I wish to express my deepest thanks to both Prof. Shkel and Anuj for their supervision throughout the semester, as taming this difficult paper alone would have probably proven to be an impossible feat. Beyond that, this project was, for me, the introduction I wanted to the world of information theory research. I am very grateful for this early opportunity to step one foot inside of it and I am looking forward to the next one.

A Useful proof ideas

A.1 Inverse and non-decreasingness

Let $f : \mathbb{R} \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$ be bijective functions. Then $f \leq g \implies f^{-1} \geq g^{-1}$

A.2 Comparing functions through their derivatives

Let $f, g : \mathbb{R}_+ \to \mathbb{R}$ be differentiable functions such that $f(0) = g(0)$. Then if $f' \geq g'$, then $f \geq g$.

A.3 Expectation as quantile function integral

Let F be a cdf, then $\mathbb{E}(F) = \int_0^1 F^{-1}(\gamma)d\gamma$.

B Misc properties of Kolmogorov infinite divisibility

We include here some classical properties of infinitely divisible distributions for completeness, with proofs in [13]. Hopefully

B.1 Summands converge to zero in probability

Given a cdf F infinitely divisible by F_n for all n, then $Y_n \sim F_n \xrightarrow{\mathbb{P}} 0$.

B.2 Equivalent definitions of infinite divisibility, link with Levy process

Let $Y \sim F$. The following statements are equivalent:

1. F is infinitely divisible.
2. $\exists \{Y_{n,j}\}_{n\geq 1, r_n \geq j \geq 1}$ i.i.d such that $r_n \overset{n\to\infty}{\to} \infty$ and $Y_{n,1} + \ldots + Y_{n,r_n} \xrightarrow{d} Y$.
3. $Y \sim X_1$ for some Levy process X.

C Spectral negative binomial random variable

Let $r, a, b \in \mathbb{N}$ and $p \in (0, 1]$. We will say $X \sim \text{SNB}(r, p, a, b)$ when

$$p_X(x) = \frac{p^r}{a}\left(\frac{1-p}{b}\right)^k$$

for $x \in \mathbb{N}$ such that

$$\sum_{i=0}^{k-1} s_i < x < \sum_{i=0}^{k} s_i, \;\; s_k = \binom{k+r-1}{r-1} ab^k$$

The number of x such that $p_X(x) = \frac{p^r}{a}\left(\frac{1-p}{b}\right)^k$ is s_k. Hence:

$$\begin{aligned}\mathbb{P}\left(p_X(X) = \frac{p^r}{a}\left(\frac{1-p}{b}\right)^k\right) &= s_k \frac{p^r}{a}\left(\frac{1-p}{b}\right)^k \\ &= \binom{k+r-1}{r-1} p^r (1-p)^k\end{aligned}$$

This is the pmf of the $\text{NegBin}(r, p)$ distribution. At the information spectrum level, we have:

$$\iota_X(X) \sim \log\left(\frac{p^r}{a}\right) + K \log\left(\frac{1-p}{b}\right), \;\; K \sim \text{NegBin}(r, p)$$

Note that the $\text{NegBin}(r, p)$ distribution is infinitely divisible, thus SNB random variables are SID. Instances of SNB random variables:

- Discrete uniform over a set of cardinality a is $\text{SNB}(1, 1, a, 1)$
- $\text{Geom}(p) = \text{SNB}(1, p, 1, 1)$

References

[1] C. E. Shannon, "A mathematical theory of communication," *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.

[2] C. T. Li, "Infinite divisibility of information," *CoRR*, vol. abs/2008.06092, 2020. [Online]. Available: https://arxiv.org/abs/2008.06092

[3] A. N. Kolmogorov, "Two uniform limit theorems for sums of independent random variables," *Theory of Probability & Its Applications*, vol. 1, no. 4, pp. 384–394, 1956. [Online]. Available: https://doi.org/10.1137/1101030

[4] C. T. Li, "Efficient approximate minimum entropy coupling of multiple probability distributions," *IEEE Transactions on Information Theory*, vol. 67, no. 8, pp. 5259–5268, 2021.

[5] Y. Y. Shkel and A. Kumar Yadav, "Information spectrum converse for minimum entropy couplings and functional representations," in *2023 IEEE International Symposium on Information Theory (ISIT)*, 2023, pp. 66–71.

[6] F. Cicalese, L. Gargano, and U. Vaccaro, "Minimum-entropy couplings and their applications," *IEEE Transactions on Information Theory*, vol. 65, no. 6, pp. 3436–3451, 2019.

[7] C. T. Li and A. E. Gamal, "Strong functional representation lemma and applications to coding theorems," *IEEE Transactions on Information Theory*, vol. 64, no. 11, pp. 6967–6978, 2018.

[8] F. Cicalese, L. Gargano, and U. Vaccaro, "Approximating probability distributions with short vectors, via information theoretic distance measures," in *2016 IEEE International Symposium on Information Theory (ISIT)*, 2016, pp. 1138–1142.

[9] E. Wolfstetter, "Stochastic dominance: Theory and applications," 02 1970.

[10] T. M. Cover and J. A. Thomas, *Elements of Information Theory 2nd Edition (Wiley Series in Telecommunications and Signal Processing), page 29.* Wiley-Interscience, July 2006.

[11] T. V. Arak, "An improvement of the lower bound for the rate of convergence in kolmogorov's uniform limit theorem," *Theory of Probability & Its Applications*, vol. 27, no. 4, pp. 826–832, 1983. [Online]. Available: https://doi.org/10.1137/1127090

[12] F. Steutel, "Infinite divisibility in theory and practice," *Scandinavian Journal of Statistics*, vol. 6, pp. 57–64, 1979.

[13] J. W. Pitman, "Levy process and infinitely divisible law," *Probability Theory*, 2003.

-