이학석사 학위논문

On the iterated image size of random functions

(랜덤 함수의 반복 이미지 크기)

2013년 11월

서울대학교 대학원
수리과학부
배 경 용

On the iterated image size of random functions

(랜덤 함수의 반복 이미지 크기)

지도교수 이 인 석

이 논문을 이학석사 학위논문으로 제출함

2013년 11월

서울대학교 대학원
수리과학부
배 경 용

배경용의 이학석사 학위논문을 인준함

2013년 11월

위 원 장 _____ ㊞

부위원장 _____ ㊞

위 원 _____ ㊞

On the iterated image size of random functions

by

Kyung Yong Bae

A DISSERTATION

Submitted to the faculty of the Graduate School
in partial fulfillment of the requirements
for the degree Master of Science
in the Department of Mathematics
Seoul National University
November, 2013

Abstract

Given a positive integer m, we denote the set $\{1, 2, \ldots, m\}$ by $[m]$. The image size of a function $f : [m] \to [n]$ is the number of elements of $f([m])$ and denoted by $|f([m])|$ or $|\mathrm{Im}\, f|$. By a random function from $[m]$ to $[n]$, we mean a function $f : [m] \to [n]$ which will be selected uniformly from the set of all functions from $[m]$ to $[n]$. Here, we emphasize that a random function is not a function which have been already selected. In this article, we compute the expectation of image size of double iterated random function.

Key Words : random function, the expectation of image size

Student Number : 2002-23256

Contents

1. Introduction

We introduce a paper 'A Comparison of Cryptanalytic Tradeoff Algorithms by J. Hong and S. Moon.' In that paper, we have found an interesting fact. <u>This introdution is the preprint of Appendix B of that paper</u>. We add our supplementary thinking to the preprint.

Cryptographers define a random function to be a function $f : [m] \rightarrow [n]$ which assigns independent and random values $f(x) \in [n]$ to all $x \in [m]$. This definition does not define any true function. Instead, the definition provides a process by which a specfic function may be constructed.

We define a random function $f : [m] \rightarrow [n]$ is a function which is chosen uniformly at random from the set of all functions from $[m]$ to $[n]$. Once more, we emphasize that this definition does not define a specific function which has been already selected. Two definitions of random function are same because the probability of being constructed or being selected is $\frac{1}{n^m}$.

Through construction approach, let's compute the expectation of the image size of a random function $f : [m] \rightarrow [n]$. First, to $1 \in [m]$, we assign a random point of $[n]$, second, to $2 \in [m]$, we assign a random point of $[n]$, ... and lastly to $m \in [m]$, we assign a random point of $[n]$. Suppose that, after a random point of $[n]$ has been assigned to k of $[m]$, the ratio of points among $[n]$ that remain as non-images is expected to be r_k. It is clear that $r_1 = \frac{n-1}{n} = \left(1 - \frac{1}{n}\right)$. After k-th process, if we assign a touched element of $[n]$ to $k+1 \in [m]$, then $r_{k+1} = r_k$. Otherwise, $r_{k+1} = r_k - \frac{1}{n}$. We get $r_{k+1} = (1 - r_k)r_k + r_k\left(r_k - \frac{1}{n}\right) = \left(1 - \frac{1}{n}\right)r_k$. Hence, $r_m = \left(1 - \frac{1}{n}\right)^m$ and the ratio of the image points among $[n]$ is expected to be $\left(1 - \left(1 - \frac{1}{n}\right)^m\right)$.

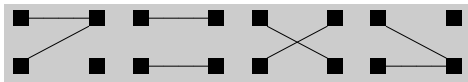
We conclude that "If one constructs a function $f : [m] \rightarrow [n]$ by assigning

independently and randomly chosen elements of $[n]$ to each input element of $[m]$, then the image size of the resulting function is expected to be $n\left(1-\left(1-\frac{1}{n}\right)^m\right)$."

If $m = n$, then we get the following statement.

Statement 1. (Correct)
Let $f : [n] \to [n]$ be a random function.
The image size $|f([n])|$ is expected to be $\alpha_1 = n\left(1-\left(1-\frac{1}{n}\right)^n\right)$.

For example, consider the set of all functions $f : [2] \to [2]$. This set can be visualized as follows.



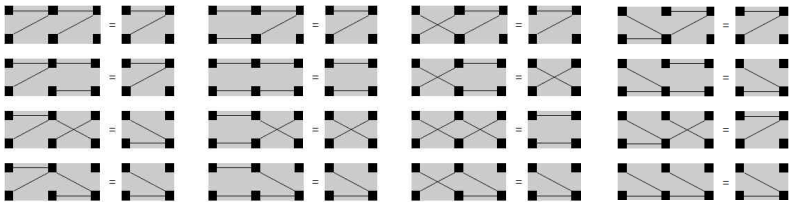
The image size expectation is $\mathrm{E}_f[|f([2])|] = 1 \cdot \frac{2}{4} + 2 \cdot \frac{2}{4} = \frac{3}{2}$, and this is also identical to the value $\alpha_1 = 2\left(1-\left(1-\frac{1}{2}\right)^2\right) = \frac{3}{2}$, computed according to statement 1. Hence, statement 1 holds exactly true for $n = 2$.

Now, we consider the following statement.

Statement 2. (Incorrect)
Let $f : [n] \to [n]$ be a random function.
Let $g : [n] \to [n]$ be a random function.
(1) The image size $|(f \circ f)([n])|$ is expected to be $\alpha_2 = n\left(1-\left(1-\frac{1}{n}\right)^{\alpha_1}\right)$.
(2) The image size $|(g \circ f)([n])|$ is expected to be $\alpha_2 = n\left(1-\left(1-\frac{1}{n}\right)^{\alpha_1}\right)$.

The claim that statement 1 implies statement 2 is acceptable in the realm of cryptology. Also, we know from experience that statement 2 works quite well in predicting the behavior of iterations done with specific functions. But there is a logical gap in statement 2. We return to the situation of $f : [2] \to [2]$. The set of all possible double iterations of the four functions can be

visualized as follows.



The expected image sizes are thus $\mathrm{E}_f[\,|(f \circ f)([2])|\,] = 1 \cdot \frac{2}{4} + 2 \cdot \frac{2}{4} = 1.5$, $\mathrm{E}_{g,f}[\,|(g \circ f)([2])|\,] = 1 \cdot \frac{12}{16} + 2 \cdot \frac{4}{16} = 1.25$ and the corresponding value, as claimed by statement 2 is $2\left(1 - \left(1 - \frac{1}{2}\right)^{\frac{3}{2}}\right) \fallingdotseq 1.293$. Hence, it is clear that statement 2 cannot be true, at least in the strict sense.

In TMTO, for sufficiently large n, "statement 1 implies statement 2" is believed as follows.

$$\mathrm{E}_f[\,|f([n])|\,] \fallingdotseq n\left(1 - \left(\frac{1}{e}\right)\right) = \alpha_1.$$

$$\mathrm{E}_{g,f}[\,|(g \circ f)([n])|\,] \fallingdotseq n\left(1 - \left(\frac{1}{e}\right)^{1-\left(\frac{1}{e}\right)}\right) = \alpha_2.$$

$$\mathrm{E}_f[\,|(f \circ f)([n])|\,] \fallingdotseq n\left(1 - \left(\frac{1}{e}\right)^{1-\left(\frac{1}{e}\right)}\right).$$

$$\frac{\alpha_2}{n} = 1 - \left(\frac{1}{e}\right)^{\frac{\alpha_1}{n}}.$$

We want to verify this belief. Thus, we embark on counting as if we are little kids counting pebbles one by one because we want to find the exact expectation of image size of f, $g \circ f$, $f \circ f$. After all, there is nothing more accurate than counting itself. Although much efforts were exerted, we could only count image size of f and $g \circ f$. Counting $f \circ f$ was extremely hard as well. We wish to launch on further research to figure out $f \circ f$.

2. Previous works

Let n and k be integers such that $1 \le n$ and $0 \le k \le n$. The binomial coefficient $\binom{n}{k}$ is the number of subsets of k objects of the set $[n]$ and the Stirling number $\left\{ {n \atop k} \right\}$ of the second kind is the number of partitions of $[n]$ into nonempty k classes. We define $\left\langle {n \atop k} \right\rangle = k!\left\{ {n \atop k} \right\}$. Clearly, $\left\langle {n \atop k} \right\rangle$ means the number of surjective functions from $[n]$ to $[k]$. Put $\binom{n}{0} = 1$, and $\left\langle {n \atop 0} \right\rangle = 0$. When $n < k$, we put $\binom{n}{k} = 0$ and $\left\langle {n \atop k} \right\rangle = 0$. Also, We put $\binom{0}{0} = 1$, $\binom{0}{1} = 0$, $\binom{0}{2} = 0$, ... and $\left\langle {0 \atop 0} \right\rangle = 1$, $\left\langle {0 \atop 1} \right\rangle = 0$, $\left\langle {0 \atop 2} \right\rangle = 0$, ...

We state two familiar facts that will be used throughout this article.

Lemma 1. For a real number a, $\lim_{n \to \infty} \left(1 - \frac{a}{n}\right)^n = \left(\frac{1}{e}\right)^a$.

Proof. See p. 43 in [1]. □

Lemma 2. For two sequences $\{a_n\}$, $\{b_n\}$ of real numbers,

(1) $\left(\sum_{n=0}^{\infty} \frac{a_n}{n!} x^n\right)\left(\sum_{n=0}^{\infty} \frac{b_n}{n!} x^n\right) = \sum_{n=0}^{\infty} \frac{1}{n!}\left(\sum_{k=0}^{n} \binom{n}{k} a_k b_{n-k}\right) x^n$.

(2) $\left(\sum_{n=j}^{\infty} \frac{a_n}{n!} x^n\right) e^x = \left(\sum_{n=j}^{\infty} \frac{a_n}{n!} x^n\right)\left(\sum_{n=0}^{\infty} \frac{1}{n!} x^n\right) = \sum_{n=j}^{\infty} \frac{1}{n!}\left(\sum_{k=j}^{n} \binom{n}{k} a_k\right) x^n$.

Proof. See p. 41 in [2]. □

From now, we prove several identities by using generating functions. These identities are necessary to counting image sizes.

Lemma 3.

If $1 \le m \le n$, $\sum_{k=1}^{m} \binom{n}{k} \left\langle {m \atop k} \right\rangle = n^m$ and if $1 \le n \le m$, $\sum_{k=1}^{n} \binom{n}{k} \left\langle {m \atop k} \right\rangle = n^m$.

Proof. Above sums mean the number of all functions from $[m]$ to $[n]$. □

Lemma 4.

(1) $\sum_{k=i}^{n}\binom{k}{i}\binom{n}{k}x^{k-i}=\binom{n}{i}(x+1)^{n-i}$.

(2) $\sum_{k=1}^{n}k\binom{n}{k}=n2^{n-1}$. (Put $i=1$ and $x=1$ in (1))

Proof. Let's differentiate $\sum_{k=0}^{n}\binom{n}{k}x^k=(x+1)^n$ with respect to x several times.

$$n(x+1)^{n-1}=\sum_{k=1}^{n}\binom{n}{k}kx^{k-1}.$$

$$n(n-1)(x+1)^{n-2}=\sum_{k=2}^{n}\binom{n}{k}k(k-1)x^{k-2}.$$

$\vdots$

$$n(n-1)\cdots(n-(i-1))(x+1)^{n-i}=\sum_{k=i}^{n}\binom{n}{k}k(k-1)\ldots(k-(i-1))x^{k-i}.$$

By dividing last eqation by $i!$, we get the following equation.

$$\frac{n(n-1)\cdots(n-(i-1))}{i!}(x+1)^{n-i}=\sum_{k=i}^{n}\binom{n}{k}\frac{k(k-1)\ldots(k-(i-1))}{i!}x^{k-i}.$$

Cosequently, $\sum_{k=i}^{n}\binom{n}{k}\binom{k}{i}x^{k-i}=\binom{n}{i}(x+1)^{n-i}$. □

The proof of the next lemma is took longer than we thought it would. Maybe, there is a way to short-circuit.

Lemma 5. Let $1\le m\le n$. Fix $1\le j\le m$.

(1) $\sum_{k=j}^{m}\binom{k}{j}\binom{n}{k}\left\langle{m \atop k}\right\rangle=\sum_{k=0}^{j}(-1)^k\binom{j}{k}\binom{n}{j}(n-k)^m$.

(2) $\sum_{k=1}^{m}k\binom{n}{k}\left\langle{m \atop k}\right\rangle=n^{m+1}-n(n-1)^m$. (Put $j=1$ in (1).)

Proof. Take a sequence $a_0=0$, $a_1=\left\langle{m \atop 1}\right\rangle$, ..., $a_m=\left\langle{m \atop m}\right\rangle$, $a_{m+1}=0$, ...

Let $f(x)=\sum_{i=0}^{\infty}\frac{a_i}{i!}x^i$ and $h(x)=f(x)e^x$.

$$h(x) = f(x)e^x$$

$$= \left(\sum_{i=0}^{\infty}\frac{a_i}{i!}x^i\right)\left(\sum_{i=0}^{\infty}\frac{1}{i!}x^i\right)$$

by Lemma 2−(2),

$$= \sum_{i=1}^{\infty}\frac{1}{i!}\left(\sum_{k=0}^{i}\binom{i}{k}a_k\right)x^i$$

$$= \frac{1}{1!}\left(\sum_{k=1}^{1}\binom{1}{k}\left\langle{m \atop k}\right\rangle\right)x^1 + \frac{1}{2!}\left(\sum_{k=1}^{2}\binom{2}{k}\left\langle{m \atop k}\right\rangle\right)x^2 + \ldots + \frac{1}{m!}\left(\sum_{k=1}^{m}\binom{m}{k}\left\langle{m \atop k}\right\rangle\right)x^m$$

$$+ \frac{1}{(m+1)!}\left(\sum_{k=1}^{m}\binom{m+1}{k}\left\langle{m \atop k}\right\rangle\right)x^{m+1} + \ldots + \frac{1}{n!}\left(\sum_{k=1}^{m}\binom{n}{k}\left\langle{m \atop k}\right\rangle\right)x^n$$

$$+ \frac{1}{(n+1)!}\left(\sum_{k=1}^{m}\binom{n+1}{k}\left\langle{m \atop k}\right\rangle\right)x^{n+1} + \ldots$$

by lemma 3,

$$= \frac{1}{1!}1^m x^1 + \frac{1}{2!}2^m x^2 + \ldots + \frac{1}{m!}m^m x^m + \frac{1}{(m+1)!}(m+1)^m x^{m+1} + \ldots$$

$$+ \frac{1}{n!}n^m x^n + \frac{1}{(n+1)!}(n+1)^m x^{n+1} + \ldots$$

$$= \sum_{i=1}^{\infty}\frac{i^m}{i!}x^i = \sum_{i=0}^{\infty}\frac{i^m}{i!}x^i. \qquad \text{(A)}$$

We differentiate $h(x) = \sum_{i=1}^{\infty}\frac{i^m}{i!}x^i$ several times.

$$h^{(1)}(x) = \sum_{i=1}^{\infty}\frac{i^m}{i!}\,i\,x^{i-1}.$$

$$h^{(2)}(x) = \sum_{i=2}^{\infty}\frac{i^m}{i!}\,i(i-1)x^{i-2}.$$

$$\vdots$$

$$h^{(k)}(x) = \sum_{i=k}^{\infty}\frac{i^m}{i!}\,\frac{i!}{(i-k)!}x^{i-k} = \sum_{i=k}^{\infty}\frac{i^m}{(i-k)!}x^{i-k}. \qquad \text{(B)}$$

We differentiate $f(x) = e^{-x}h(x)$ and $f(x) = \sum_{i=0}^{\infty}\frac{a_i}{i!}x^i$ j times respectively and compute $e^x x^j f^{(j)}$ in two different ways and compare their coefficients of x^n.

$f(x)=e^{-x}h(x).$

$f^{(1)}(x)=e^{-x}\left(h^{(1)}(x)-h(x)\right).$

$f^{(2)}(x)=e^{-x}\left(h^{(2)}(x)-2h^{(1)}(x)+h(x)\right).$

$f^{(3)}(x)=e^{-x}\left(h^{(3)}(x)-3h^{(2)}(x)+3h^{(1)}(x)-h(x)\right).$

$\vdots$

$$f^{(j)}(x)=e^{-x}\left(\sum_{k=0}^{j-1}(-1)^k\binom{j}{k}h^{(j-k)}(x)+(-1)^j h(x)\right).$$

$$e^x x^j f^{(j)}=x^j\left(\sum_{k=0}^{j-1}(-1)^k\binom{j}{k}h^{(j-k)}(x)+(-1)^j h(x)\right)$$

by (A) and (B),

$$=x^j\left(\sum_{k=0}^{j-1}(-1)^k\binom{j}{k}\sum_{i=j-k}^{\infty}\frac{i^m}{(i-(j-k))!}x^{i-(j-k)}+(-1)^j\sum_{i=0}^{\infty}\frac{i^m}{i!}x^i\right)$$

$$=\sum_{k=0}^{j-1}(-1)^k\binom{j}{k}\sum_{i=j-k}^{\infty}\frac{i^m}{(i-(j-k))!}x^{i+k}+(-1)^j\sum_{i=0}^{\infty}\frac{i^m}{i!}x^{i+j}.$$

The coefficient of x^n $=\left(\sum_{k=0}^{j-1}(-1)^k\binom{j}{k}\frac{(n-k)^m}{((n-k)-(j-k))!}\right)+(-1)^j\frac{(n-j)^m}{(n-j)!}$

$$=\left(\sum_{k=0}^{j-1}(-1)^k\binom{j}{k}\frac{(n-k)^m}{(n-j)!}\right)+(-1)^j\frac{(n-j)^m}{(n-j)!}$$

$$=\sum_{k=0}^{j}(-1)^k\binom{j}{k}\frac{(n-k)^m}{(n-j)!}. \qquad \text{(C)}$$

$f(x)=\sum_{i=0}^{\infty}\frac{a_i}{i!}x^i.$

$f^{(1)}(x)=\sum_{i=1}^{\infty}\frac{a_i}{(i-1)!}x^{i-1}.$

$f^{(2)}(x)=\sum_{i=2}^{\infty}\frac{a_i}{(i-2)!}x^{i-2}.$

$f^{(3)}(x)=\sum_{i=3}^{\infty}\frac{a_i}{(i-3)!}x^{i-3}.$

$\vdots$

$f^{(j)}(x)=\sum_{i=j}^{\infty}\frac{a_i}{(i-j)!}x^{i-j}.$

$$x^j f^{(j)} = x^j \sum_{i=j}^{\infty} \frac{a_i}{(i-j)!} x^{i-j} = \sum_{i=j}^{\infty} \frac{a_i}{(i-j)!} x^i = \sum_{i=j}^{\infty} \frac{1}{i!} \frac{i!\, a_i}{(i-j)!} x^i.$$

By lemma 2−(2), $e^x x^j f^{(j)} = \left(\sum_{i=j}^{\infty} \frac{1}{i!} \frac{i!\, a_i}{(i-j)!} x^i \right) e^x = \sum_{i=j}^{\infty} \frac{1}{i!} \left(\sum_{k=j}^{i} \binom{i}{k} \frac{k!\, a_k}{(k-j)!} \right) x^i$.

The coefficient of $x^n = \frac{1}{n!} \left(\sum_{k=j}^{n} \binom{n}{k} \frac{k!\, a_k}{(k-j)!} \right)$

because $a_{m+1} = 0$, $a_{m+2} = 0$, ..., $a_n = 0$,

$$= \frac{1}{n!} \left(\sum_{k=j}^{m} \binom{n}{k} \frac{k!\, a_k}{(k-j)!} \right) = \frac{1}{n!} \sum_{k=j}^{m} \frac{k!}{(k-j)!} \binom{n}{k} \left\langle {m \atop k} \right\rangle. \quad \text{(D)}$$

By (C) and (D), $\sum_{k=j}^{m} \frac{k!}{(k-j)!} \binom{n}{k} \left\langle {m \atop k} \right\rangle = n! \sum_{k=0}^{j} (-1)^k \binom{j}{k} \frac{(n-k)^m}{(n-j)!}$

$$= \sum_{k=0}^{j} (-1)^k \binom{j}{k} \frac{n!}{(n-j)!} (n-k)^m.$$

By dividing above identity by $j!$, the following equation holds.

$$\sum_{k=j}^{m} \frac{k!}{(k-j)!\, j!} \binom{n}{k} \left\langle {m \atop k} \right\rangle = \sum_{k=0}^{j} (-1)^k \binom{j}{k} \frac{n!}{(n-j)!\, j!} (n-k)^m.$$

Thus we conclude that $\sum_{k=j}^{m} \binom{k}{j} \binom{n}{k} \left\langle {m \atop k} \right\rangle = \sum_{k=0}^{j} (-1)^k \binom{j}{k} \binom{n}{j} (n-k)^m$. □

The next lemma is necessary to computing expectation and variance of double iterated functions.

Lemma 6. Given a real number β,

$$\sum_{a=1}^{n} \left(1 - \frac{\beta}{n}\right)^a \binom{n}{a} \left\langle {n \atop a} \right\rangle = n^n \left(1 - \frac{\beta}{n}\right)^n \sum_{a=0}^{n} \binom{n}{a} \left(\frac{\beta}{n-\beta}\right)^a \left(1 - \frac{a}{n}\right)^n.$$

Proof. $\sum_{a=1}^{n} \left(1 - \frac{\beta}{n}\right)^a \binom{n}{a} \left\langle {n \atop a} \right\rangle = \sum_{a=1}^{n} \binom{n}{a} \left\langle {n \atop a} \right\rangle \sum_{j=0}^{a} \binom{a}{j} \left(-\frac{\beta}{n}\right)^j$

$$= \sum_{a=1}^{n}\binom{n}{a}\left\langle {n \atop a} \right\rangle + \sum_{j=1}^{n}\sum_{a=j}^{n}\binom{n}{a}\left\langle {n \atop a} \right\rangle\binom{a}{j}\left(-\frac{\beta}{n}\right)^{j}$$

$$= n^{n} + \sum_{j=1}^{n}\left(-\frac{\beta}{n}\right)^{j}\sum_{a=j}^{n}\binom{a}{j}\binom{n}{a}\left\langle {n \atop a} \right\rangle$$

by lemma 5−(1),

$$= n^{n} + \sum_{j=1}^{n}\left(-\frac{\beta}{n}\right)^{j}\sum_{a=0}^{j}(-1)^{a}\binom{j}{a}\binom{n}{j}(n-a)^{n}$$

$$= n^{n} + \sum_{j=1}^{n}\binom{n}{j}\left(-\frac{\beta}{n}\right)^{j}\sum_{a=0}^{j}(-1)^{a}\binom{j}{a}(n-a)^{n}$$

$$= n^{n} + n^{n}\sum_{j=1}^{n}\binom{n}{j}\left(-\frac{\beta}{n}\right)^{j}\sum_{a=0}^{j}(-1)^{a}\binom{j}{a}\left(1-\frac{a}{n}\right)^{n}$$

$$= n^{n} + n^{n}\left[\sum_{j=1}^{n}\binom{n}{j}\left(-\frac{\beta}{n}\right)^{j} + \sum_{a=1}^{n}\sum_{j=a}^{n}\binom{n}{j}\left(-\frac{\beta}{n}\right)^{j}(-1)^{a}\binom{j}{a}\left(1-\frac{a}{n}\right)^{n}\right]$$

$$= n^{n} + n^{n}\left[\sum_{j=1}^{n}\binom{n}{j}\left(-\frac{\beta}{n}\right)^{j} + \sum_{a=1}^{n}(-1)^{a}\left(1-\frac{a}{n}\right)^{n}\sum_{j=a}^{n}\binom{j}{a}\binom{n}{j}\left(-\frac{\beta}{n}\right)^{j}\right]$$

$$= n^{n} + n^{n}\left[\sum_{j=1}^{n}\binom{n}{j}\left(-\frac{\beta}{n}\right)^{j} + \sum_{a=1}^{n}(-1)^{a}\left(-\frac{\beta}{n}\right)^{a}\left(1-\frac{a}{n}\right)^{n}\sum_{j=a}^{n}\binom{j}{a}\binom{n}{j}\left(-\frac{\beta}{n}\right)^{j-a}\right]$$

by lemma 4−(1),

$$= n^{n} + n^{n}\left[\sum_{j=1}^{n}\binom{n}{j}\left(-\frac{\beta}{n}\right)^{j} + \sum_{a=1}^{n}(-1)^{a}\left(-\frac{\beta}{n}\right)^{a}\left(1-\frac{a}{n}\right)^{n}\binom{n}{a}\left(1-\frac{\beta}{n}\right)^{n-a}\right]$$

$$= n^{n} + n^{n}\left[\left(1-\frac{\beta}{n}\right)^{n} - 1 + \left(1-\frac{\beta}{n}\right)^{n}\sum_{a=1}^{n}\left(\frac{\beta}{n}\right)^{a}\left(1-\frac{a}{n}\right)^{n}\binom{n}{a}\left(1-\frac{\beta}{n}\right)^{-a}\right]$$

$$= n^{n} + n^{n}\left[\left(1-\frac{\beta}{n}\right)^{n} - 1 + \left(1-\frac{\beta}{n}\right)^{n}\sum_{a=1}^{n}\binom{n}{a}\left(\frac{n}{\beta}-1\right)^{-a}\left(1-\frac{a}{n}\right)^{n}\right]$$

$$= n^{n} + n^{n}\left[\left(1-\frac{\beta}{n}\right)^{n} - 1 + \left(1-\frac{\beta}{n}\right)^{n}\sum_{a=1}^{n}\binom{n}{a}\left(\frac{\beta}{n-\beta}\right)^{a}\left(1-\frac{a}{n}\right)^{n}\right]$$

$$= n^{n}\left(1-\frac{\beta}{n}\right)^{n}\sum_{a=0}^{n}\binom{n}{a}\left(\frac{\beta}{n-\beta}\right)^{a}\left(1-\frac{a}{n}\right)^{n}. \qquad \square$$

The following lemma is necessary to computing expectation of inverse image size.

Lemma 7. Given $1 \le n$, $a \ne 0$ and $b \ne 0$, $\sum_{k=0}^{n} k\binom{n}{k}a^k b^{n-k} = an(a+b)^{n-1}$.

Proof. Let $f(x) = \sum_{i=0}^{\infty} \frac{a^i b^{n-i}}{i!} x^i$ and $h(x) = f(x)e^x$.

By lemma 2, $h(x) = f(x)e^x = \sum_{i=0}^{\infty} \frac{1}{i!}\left(\sum_{k=0}^{i}\binom{i}{k}a^k b^{n-k}\right)x^i$.

Like the proof of lemma 5, we express $e^x x f^{/}(x)$ in two different ways and compare coefficients of x^n.

$f(x) = e^{-x}h(x)$. $f^{/}(x) = e^{-x}\left(h^{/}(x) - h(x)\right)$.

$$\begin{aligned} e^x x f^{/} &= x\left(h^{/}(x) - h(x)\right) \\ &= x\left(\sum_{i=1}^{\infty}\frac{i}{i!}\left(\sum_{k=0}^{i}\binom{i}{k}a^k b^{n-k}\right)x^{i-1} - \sum_{i=0}^{\infty}\frac{1}{i!}\left(\sum_{k=0}^{i}\binom{i}{k}a^k b^{n-k}\right)x^i\right). \end{aligned}$$

The coefficient of $x^n = \frac{n}{n!}\left(\sum_{k=0}^{n}\binom{n}{k}a^k b^{n-k}\right) - \frac{1}{(n-1)!}\left(\sum_{k=0}^{n-1}\binom{n-1}{k}a^k b^{n-k}\right)$. (E)

$f(x) = \sum_{i=0}^{\infty}\frac{a^i b^{n-i}}{i!}x^i$. $xf^{/} = \sum_{i=0}^{\infty}\frac{i\,a^i b^{n-i}}{i!}x^i$. $e^x x f^{/} = \sum_{i=0}^{\infty}\frac{1}{i!}\left(\sum_{k=0}^{i}\binom{i}{k}k a^k b^{n-k}\right)x^i$.

The coefficient of $x^n = \frac{1}{n!}\sum_{k=0}^{n} k\binom{n}{k}a^k b^{n-k}$. (F)

By (E) and (F), $\sum_{k=0}^{n} k\binom{n}{k}a^k b^{n-k} = n\left(\sum_{k=0}^{n}\binom{n}{k}a^k b^{n-k}\right) - n\left(\sum_{k=0}^{n-1}\binom{n-1}{k}a^k b^{n-k}\right)$.

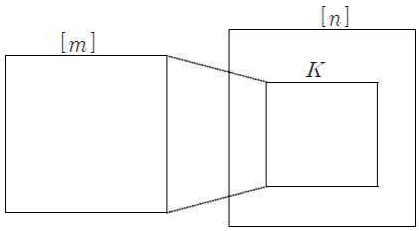
$$\begin{aligned} \text{Thus, } \sum_{k=0}^{n} k\binom{n}{k}a^k b^{n-k} &= n(a+b)^n - nb(a+b)^{n-1} \\ &= n(a+b)^{n-1}((a+b) - b) = an(a+b)^{n-1}. \end{aligned}$$ □

3. Expectation of image size of random function

We try to compute the expectation of image size of random function by counting all image sizes of all functions, that is, through selection approach.

Definition 8. A random function $f : [m] \to [n]$ is a function which will be chosen uniformly at random from the set of all functions from $[m]$ to $[n]$.

Let $1 \le m \le n$, consider the set $\mathscr{A}$ of all functions from $[m]$ to $[n]$. Let's choose all functions having image size k. Note that $1 \le k \le m$. These functions are surjective functions from $[m]$ to a subset K, having k elements, of $[n]$.



Thus, the number of functions having image size k is $\binom{n}{k}\left\langle {m \atop k} \right\rangle$. (G)

Lemma 9. If $1 \le m \le n$. The image size of a random function $f : [m] \to [n]$ is expected to be $n\left(1-\left(1-\frac{1}{n}\right)^m\right)$.

Proof. By (G) and lemma 5-(2),

$$\frac{\sum_{f:[m]\to[n]} |\mathrm{Im}\, f|}{n^m} = \frac{\sum_{k=1}^{m} k\binom{n}{k}\left\langle {m \atop k} \right\rangle}{n^m} = \frac{n^{m+1} - n(n-1)^m}{n^m} = n\left(1-\left(1-\frac{1}{n}\right)^m\right). \square$$

The proportional image size of a function $f : [m] \to [n]$ is defined as $\frac{|\mathrm{Im}\, f|}{n}$.

By lemma 9, we see that the proportional image size of a random function $f : [m] \to [n]$ is expected to be $\dfrac{\sum_{f:[m]\to[n]} \frac{|\mathrm{Im}\, f|}{n}}{n^m} = \left(1-\left(1-\frac{1}{n}\right)^m\right)$.

Let's determine the limit of $\left(1-\left(1-\frac{1}{n}\right)^m\right)$ as $m,\ n \to \infty$ with $\frac{m}{n}=r$ fixed. This will be the asymptotic proportional image size of a random function.

Lemma 10. Fix $0 < r \le 1$, If $\frac{m}{n}=r$, then $\lim_{n\to\infty}\left(1-\left(1-\frac{1}{n}\right)^m\right) = 1-\left(\frac{1}{e}\right)^r$.

Proof. $\lim_{n\to\infty} 1-\left(1-\frac{1}{n}\right)^m = \lim_{n\to\infty} 1-\left(\left(1-\frac{1}{n}\right)^n\right)^r = 1-\left(\frac{1}{e}\right)^r$ by lemma 1. □

Now, consider proportional image sizes of all functions of $\mathscr{A}$ with $\frac{m}{n}=r$ fixed. Let's determine the variance for the proportional image size.

The variance is $\dfrac{\sum_{f:[m]\to[n]}\left(\frac{|\mathrm{Im}\, f|}{n}\right)^2}{n^m} - \left(\dfrac{\sum_{f:[m]\to[n]}\frac{|\mathrm{Im}\, f|}{n}}{n^m}\right)^2$

$$= \frac{\sum_{k=1}^{m}\left(\frac{k}{n}\right)^2\binom{n}{k}\left\langle {m \atop k}\right\rangle}{n^m} - \left(\frac{\sum_{k=1}^{m}\frac{k}{n}\binom{n}{k}\left\langle {m \atop k}\right\rangle}{n^m}\right)^2.$$

$$\sum_{k=2}^{m}\binom{k}{2}\binom{n}{k}\left\langle {m \atop k}\right\rangle = \sum_{k=2}^{m}\frac{k(k-1)}{2}\binom{n}{k}\left\langle {m \atop k}\right\rangle = \frac{1}{2}\left[\sum_{k=1}^{m}k^2\binom{n}{k}\left\langle {m \atop k}\right\rangle - \sum_{k=1}^{m}k\binom{n}{k}\left\langle {m \atop k}\right\rangle\right].$$

$$\sum_{k=1}^{m}k^2\binom{n}{k}\left\langle {m \atop k}\right\rangle = 2\sum_{k=2}^{m}\binom{k}{2}\binom{n}{k}\left\langle {m \atop k}\right\rangle + \sum_{k=1}^{m}k\binom{n}{k}\left\langle {m \atop k}\right\rangle$$

By lemma 5,

$$= 2\sum_{k=0}^{2}(-1)^k\binom{2}{k}\binom{n}{2}(n-k)^m + \left(n^{m+1} - n(n-1)^m\right)$$

$$= \left(n^2-n\right)\left(n^m - 2(n-1)^m + (n-2)^m\right) + \left(n^{m+1} - n(n-1)^m\right)$$

$$= n^{m+2} - (n-1)^m\left(2n^2-n\right) + (n-2)^m\left(n^2-n\right). \qquad \text{(H)}$$

$$\frac{\sum_{k=1}^{m}\left(\frac{k}{n}\right)^2\binom{n}{k}\left\langle {m \atop k}\right\rangle}{n^m} - \left(\frac{\sum_{k=1}^{m}\frac{k}{n}\binom{n}{k}\left\langle {m \atop k}\right\rangle}{n^m}\right)^2$$

$$= \frac{n^{m+2}-(n-1)^m(2n^2-n)+(n-2)^m(n^2-n)}{n^{m+2}} - \left(\frac{n^{m+1}-n(n-1)^m}{n^{m+1}}\right)^2$$

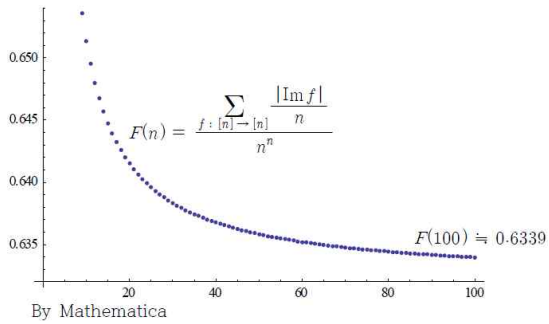
$$= 1-\left(1-\frac{1}{n}\right)^m\left(2-\frac{1}{n}\right)+\left(1-\frac{2}{n}\right)^m\left(1-\frac{1}{n}\right)-\left(1-\left(1-\frac{1}{n}\right)^m\right)^2$$

$$\to 1-2\left(\frac{1}{e}\right)^r+\left(\frac{1}{e}\right)^{2r}-\left(1-\left(\frac{1}{e}\right)^r\right)^2 = 0 \text{ by lemma 1.}$$

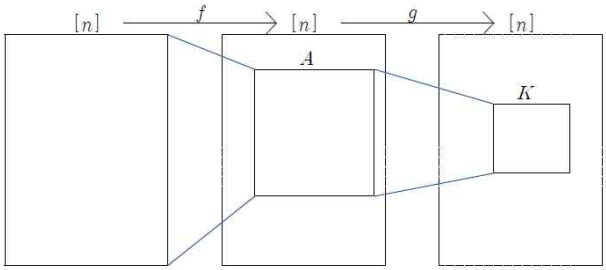
Thus, the probability distribution of the proportional image size becomes more and more concentrated at $1-\left(\frac{1}{e}\right)^r$. We think that the following theorem is believed in this background.

Theorem 11. Let $0 < r \le 1$ and $\frac{m}{n} = r$. The image size of a random function $f : [m] \to [n]$ is expected to be $n\left(1-\left(1-\frac{1}{n}\right)^m\right) \fallingdotseq n\left(1-\left(\frac{1}{e}\right)^r\right)$ if n is sufficiently large.

Corollary 12. If n is sufficiently large, then the image size of a random function $f : [n] \to [n]$ is expected to be $n\left(1-\left(1-\frac{1}{n}\right)^n\right) \fallingdotseq n\left(1-\frac{1}{e}\right) \fallingdotseq (0.6321)n$.



For $1 \leq n$, let $\mathscr{B}$ be the set of all functions from $[n]$ to $[n]$ and $\mathscr{C}$ be the set of all double iterated functions $g \circ f$ from $[n]$ to $[n]$ where f and g are elements of $\mathscr{B}$. We define two functions $g_1 \circ f_1$ and $g_2 \circ f_2$ of $\mathscr{C}$ are equal if $g_1 = g_2$ and $f_1 = f_2$.



We select all iterated functions $g \circ f$ from $\mathscr{C}$ so that their image size is k. First, we choose a subset A, having a elements, of $[n]$. Second, we choose a subset K, having k elements, of $[n]$. Third, we choose all iterated functions $g \circ f$ such that f is a surjective function from $[n]$ to A and g is a surjctive function from A to K when the domain of g is restricted to A. We can see that the number of double iterated functions having image size k is

$$\sum_{a=1}^{n} \binom{n}{a} \left\langle {n \atop a} \right\rangle \binom{n}{k} \left\langle {a \atop k} \right\rangle n^{n-a}. \qquad \text{(I)}$$

We try to compute the expectation of the image size of a iterated function $g \circ f$ of $\mathscr{C}$. Before computation, let's conjecture the expectation for sufficiently large n. By corollary 12, the image size of f is expected to be $n\left(1 - \frac{1}{e}\right)$ and by theorem 11, the image size of $g : n\left(1 - \frac{1}{e}\right) \to [n]$ is expected to be $n\left(1 - \left(\frac{1}{e}\right)^{1-\left(\frac{1}{e}\right)}\right)$. Thus, we conjecture that, for sufficiently large n, the image size of a iterated function $g \circ f$ is expected to be $n\left(1 - \left(\frac{1}{e}\right)^{1-\left(\frac{1}{e}\right)}\right)$. Now, we begin computation.

Lemma 13. Let $1 \le n$. If f and g are random functions from $[n] \to [n]$, then, the image size of a random function $g \circ f : [n] \to [n]$ is expected to be $n\left[1 - \left(1 - \frac{1}{n}\right)^n \sum_{a=0}^{n} \binom{n}{a}\left(\frac{1}{n-1}\right)^a \left(1 - \frac{a}{n}\right)^n\right]$.

Proof. By (I),

$$\frac{\sum_{\substack{f:[n]\to[n] \\ g:[n]\to[n]}} |\mathrm{Im}(g \circ f)|}{n^{2n}} = \frac{\sum_{k=1}^{n} k \sum_{a=1}^{n} \binom{n}{a}\left\langle {n \atop a} \right\rangle \binom{n}{k}\left\langle {a \atop k} \right\rangle n^{n-a}}{n^{2n}}$$

$$= \frac{\sum_{a=1}^{n} \sum_{k=1}^{a} k\binom{n}{a}\left\langle {n \atop a} \right\rangle \binom{n}{k}\left\langle {a \atop k} \right\rangle n^{n-a}}{n^{2n}}$$

$$= \frac{\sum_{a=1}^{n} \binom{n}{a}\left\langle {n \atop a} \right\rangle n^{n-a} \sum_{k=1}^{a} k\binom{n}{k}\left\langle {a \atop k} \right\rangle}{n^{2n}}$$

by lemma 5−(2),

$$= \frac{\sum_{a=1}^{n} \binom{n}{a}\left\langle {n \atop a} \right\rangle n^{n-a}\left(n^{a+1} - n(n-1)^a\right)}{n^{2n}}$$

$$= \frac{\sum_{a=1}^{n} n^{n+1}\binom{n}{a}\left\langle {n \atop a} \right\rangle - \sum_{a=1}^{n} n^{n-a} n(n-1)^a \binom{n}{a}\left\langle {n \atop a} \right\rangle}{n^{2n}}$$

$$= \frac{n^{n+1}\sum_{a=1}^{n} \binom{n}{a}\left\langle {n \atop a} \right\rangle - n^{n+1}\sum_{a=1}^{n} \left(\frac{1}{n}\right)^a (n-1)^a \binom{n}{a}\left\langle {n \atop a} \right\rangle}{n^{2n}}$$

$$= \frac{n^{n+1}n^n - n^{n+1}\sum_{a=1}^{n} \left(1 - \frac{1}{n}\right)^a \binom{n}{a}\left\langle {n \atop a} \right\rangle}{n^{2n}}$$

by lemma 6,

$$= \frac{n^{n+1}n^n - n^{n+1}n^n\left(1 - \frac{1}{n}\right)^n \sum_{a=0}^{n} \binom{n}{a}\left(\frac{1}{n-1}\right)^a \left(1 - \frac{a}{n}\right)^n}{n^{2n}}$$

$$= n\left[1 - \left(1 - \frac{1}{n}\right)^n \sum_{a=0}^{n} \binom{n}{a}\left(\frac{1}{n-1}\right)^a \left(1 - \frac{a}{n}\right)^n\right]. \qquad \square$$

By lemma 13, the proportional image size of a double iterated random function $g \circ f : [n] \to [n]$ is expected to be

$$\frac{\displaystyle\sum_{\substack{f:[n]\to[n]\\ g:[n]\to[n]}} \frac{|\mathrm{Im}(g \circ f)|}{n}}{n^{2n}} = 1 - \left(1 - \frac{1}{n}\right)^n \sum_{a=0}^{n} \binom{n}{a} \left(\frac{1}{n-1}\right)^a \left(1 - \frac{a}{n}\right)^n. \qquad \text{(J)}$$

We determine the limit of $1 - \left(1 - \frac{1}{n}\right)^n \sum_{a=0}^{n} \binom{n}{a} \left(\frac{1}{n-1}\right)^a \left(1 - \frac{a}{n}\right)^n$ as $n \to \infty$.

This limit will be the asymptotic proportional image size of a double iterated random function.

Lemma 14. $\displaystyle\lim_{n\to\infty} 1 - \left(1 - \frac{1}{n}\right)^n \sum_{a=0}^{n} \binom{n}{a} \left(\frac{1}{n-1}\right)^a \left(1 - \frac{a}{n}\right)^n = 1 - \left(\frac{1}{e}\right)^{1-\left(\frac{1}{e}\right)} \fallingdotseq 0.4685.$

Proof. By lemma 1,

$$1 - \left(1 - \frac{1}{n}\right)^n \sum_{a=0}^{n} \binom{n}{a} \left(\frac{1}{n-1}\right)^a \left(1 - \frac{a}{n}\right)^n$$

$$= 1 - \left(1 - \frac{1}{n}\right)^n \left[1 + \frac{1}{1!}\frac{n}{n-1}\left(1 - \frac{1}{n}\right)^n + \frac{1}{2!}\frac{n(n-1)}{(n-1)^2}\left(1 - \frac{2}{n}\right)^n + \frac{1}{3!}\frac{n(n-1)(n-2)}{(n-1)^3}\left(1 - \frac{3}{n}\right)^n + \right.$$

$$\left. \cdots + \frac{1}{(n-1)!}\frac{n(n-1)(n-2)\cdots(n-(n-2))}{(n-1)^{n-1}}\left(1 - \frac{n-1}{n}\right)^n\right]$$

$$\to 1 - \frac{1}{e}\left(1 + \frac{1}{1!}\left(\frac{1}{e}\right) + \frac{1}{2!}\left(\frac{1}{e}\right)^2 + \frac{1}{3!}\left(\frac{1}{e}\right)^3 + \cdots\right) = 1 - \left(\frac{1}{e}\right)e^{\left(\frac{1}{e}\right)} = 1 - \left(\frac{1}{e}\right)^{1-\left(\frac{1}{e}\right)}. \square$$

We consider proportional image sizes of all double iterated functions of $\mathscr{C}$. Let's determine the variance for their proportional image size. The variance is

$$\frac{\displaystyle\sum_{\substack{f:[n]\to[n]\\ g:[n]\to[n]}} \left(\frac{|\mathrm{Im}(g \circ f)|}{n}\right)^2}{n^{2n}} - \left(\frac{\displaystyle\sum_{\substack{f:[n]\to[n]\\ g:[n]\to[n]}} \frac{|\mathrm{Im}(g \circ f)|}{n}}{n^{2n}}\right)^2$$

$$= \frac{\displaystyle\sum_{a=1}^{n}\sum_{k=1}^{a} \left(\frac{k}{n}\right)^2 \binom{n}{a}\left\langle {n \atop a} \right\rangle \binom{n}{k}\left\langle {a \atop k} \right\rangle n^{n-a}}{n^{2n}} - \left(\frac{\displaystyle\sum_{a=1}^{n}\sum_{k=1}^{a} \frac{k}{n} \binom{n}{a}\left\langle {n \atop a} \right\rangle \binom{n}{k}\left\langle {a \atop k} \right\rangle n^{n-a}}{n^{2n}}\right)^2$$

by (J),

$$= \frac{\sum_{a=1}^{n}\binom{n}{a}\left\langle {n \atop a}\right\rangle n^{n-a}\sum_{k=1}^{a}k^2\binom{n}{k}\left\langle {a \atop k}\right\rangle}{n^{2n+2}} - \left(1-\left(1-\frac{1}{n}\right)^n\sum_{a=0}^{n}\binom{n}{a}\left(\frac{1}{n-1}\right)^a\left(1-\frac{a}{n}\right)^n\right)^2. \quad \text{(K)}$$

$$\frac{\sum_{a=1}^{n}\binom{n}{a}\left\langle {n \atop a}\right\rangle n^{n-a}\sum_{k=1}^{a}k^2\binom{n}{k}\left\langle {a \atop k}\right\rangle}{n^{2n+2}}$$

by (H),

$$= \frac{\sum_{a=1}^{n}\binom{n}{a}\left\langle {n \atop a}\right\rangle n^{n-a}\left[n^{a+2}-(n-1)^a(2n^2-n)+(n-2)^a(n^2-n)\right]}{n^{2n+2}}$$

$$= \frac{\sum_{a=1}^{n}\binom{n}{a}\left\langle {n \atop a}\right\rangle \left[n^{n+2}-n^n\left(1-\frac{1}{n}\right)^a(2n^2-n)+n^n\left(1-\frac{2}{n}\right)^a(n^2-n)\right]}{n^{2n+2}}$$

$$= \frac{\sum_{a=1}^{n}\binom{n}{a}\left\langle {n \atop a}\right\rangle}{n^n} - \frac{(2n^2-n)\sum_{a=1}^{n}\left(1-\frac{1}{n}\right)^a\binom{n}{a}\left\langle {n \atop a}\right\rangle}{n^{n+2}} + \frac{(n^2-n)\sum_{a=1}^{n}\left(1-\frac{2}{n}\right)^a\binom{n}{a}\left\langle {n \atop a}\right\rangle}{n^{n+2}}$$

by Lemma 6,

$$= \frac{\sum_{a=1}^{n}\binom{n}{a}\left\langle {n \atop a}\right\rangle}{n^n} - \frac{(2n^2-n)n^n\left(1-\frac{1}{n}\right)^n\sum_{a=0}^{n}\binom{n}{a}\left(\frac{1}{n-1}\right)^a\left(1-\frac{a}{n}\right)^n}{n^{n+2}}$$

$$+ \frac{(n^2-n)n^n\left(1-\frac{2}{n}\right)^n\sum_{a=0}^{n}\binom{n}{a}\left(\frac{2}{n-2}\right)^a\left(1-\frac{a}{n}\right)^n}{n^{n+2}}$$

$$= \frac{n^n}{n^n} - \left(2-\frac{1}{n}\right)\left(1-\frac{1}{n}\right)^n\sum_{a=0}^{n}\binom{n}{a}\left(\frac{1}{n-1}\right)^a\left(1-\frac{a}{n}\right)^n$$

$$+\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)^n\sum_{a=0}^{n}\binom{n}{a}\left(\frac{2}{n-2}\right)^a\left(1-\frac{a}{n}\right)^n$$

$$= 1 - \left(2-\frac{1}{n}\right)\left(1-\frac{1}{n}\right)^n\left[1+\frac{1}{1!}\frac{n}{n-1}\left(1-\frac{1}{n}\right)^n+\frac{1}{2!}\frac{n(n-1)}{(n-1)^2}\left(1-\frac{2}{n}\right)^n+\ldots\right.$$

$$\left.\ldots+\frac{1}{(n-1)!}\frac{n(n-1)(n-2)\ldots(n-(n-2))}{(n-1)^{n-1}}\left(1-\frac{n-1}{n}\right)^n\right]$$

$$+\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)^n\left[1+\frac{1}{1!}\frac{n}{n-2}2^1\left(1-\frac{1}{n}\right)^n+\frac{1}{2!}\frac{n(n-1)}{(n-2)^2}2^2\left(1-\frac{2}{n}\right)^n+\ldots\right.$$

$$\ldots + \frac{1}{(n-1)!}\frac{n(n-1)(n-2)\ldots(n-(n-2))}{(n-2)^{n-1}}2^{n-1}\left(1-\frac{n-1}{n}\right)^n\Bigg]$$

by lemma 1,

$$\to 1-2\frac{1}{e}\left(1+\frac{1}{1!}\left(\frac{1}{e}\right)+\frac{1}{2!}\left(\frac{1}{e}\right)^2+\frac{1}{3!}\left(\frac{1}{e}\right)^3+..\right)+\left(\frac{1}{e}\right)^2\left(1+\frac{1}{1!}\left(\frac{2}{e}\right)+\frac{1}{2!}\left(\frac{2}{e}\right)^2+\frac{1}{3!}\left(\frac{2}{e}\right)^3+..\right)$$

$$= 1-2\left(\frac{1}{e}\right)e^{\frac{1}{e}}+\left(\frac{1}{e}\right)^2 e^{\frac{2}{e}} = 1-2\left(\frac{1}{e}\right)^{1-\left(\frac{1}{e}\right)}+\left(\frac{1}{e}\right)^{2-\left(\frac{2}{e}\right)} = \left(1-\left(\frac{1}{e}\right)^{1-\left(\frac{1}{e}\right)}\right)^2.$$

Thus, $$\frac{\sum_{a=1}^{n}\binom{n}{a}\left\langle{n \atop a}\right\rangle n^{n-a}\sum_{k=1}^{a}k^2\binom{n}{k}\left\langle{a \atop k}\right\rangle}{n^{2n+2}} \to \left(1-\left(\frac{1}{e}\right)^{1-\left(\frac{1}{e}\right)}\right)^2. \qquad \text{(L)}$$

By lemma 14, (K) and (L),

$$\frac{\sum_{\substack{f:[n]\to[n]\\ g:[n]\to[n]}}\left(\frac{|\mathrm{Im}(g\circ f)|}{n}\right)^2}{n^{2n}} - \left(\frac{\sum_{\substack{f:[n]\to[n]\\ g:[n]\to[n]}}\frac{|\mathrm{Im}(g\circ f)|}{n}}{n^{2n}}\right)^2$$

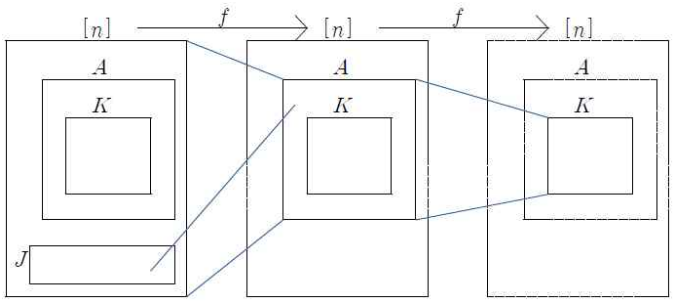
$$= \frac{\sum_{a=1}^{n}\binom{n}{a}\left\langle{n \atop a}\right\rangle n^{n-a}\sum_{k=1}^{a}k^2\binom{n}{k}\left\langle{a \atop k}\right\rangle}{n^{2n+2}} - \left(1-\left(1-\frac{1}{n}\right)^n\sum_{a=0}^{n}\binom{n}{a}\left(\frac{1}{n-1}\right)^a\left(1-\frac{a}{n}\right)^n\right)^2$$

$$\to \left(1-\left(\frac{1}{e}\right)^{1-\left(\frac{1}{e}\right)}\right)^2-\left(1-\left(\frac{1}{e}\right)^{1-\left(\frac{1}{e}\right)}\right)^2 = 0.$$

Now, we see that the limit of the variance of proportional image size for all double iterated functions $g\circ f$ of $\mathscr{C}$ is zero. Thus, their probability distribution also becomes more and more concentrated at $1-\left(\frac{1}{e}\right)^{1-\left(\frac{1}{e}\right)}$. The following theorem is based on this argument.

Theorem 15. If f and g are random functions from $[n] \to [n]$, then, the image size of a random function $g\circ f : [n] \to [n]$ is expected to be $n\left(1-\left(\frac{1}{e}\right)^{1-\left(\frac{1}{e}\right)}\right) \fallingdotseq (0.4685)n$ if n is sufficiently large.

Now, we return to the set $\mathscr{B}$ of all functions from $[n]$ to $[n]$ and consider all double iterated functions $f \circ f$ where f is an element of $\mathscr{B}$.



Let's select all functions f from $\mathscr{B}$ such that the image size of $f \circ f$ is k. First, we choose a subset A, having a elements, of $[n]$. Second, we choose a subset K, having k elements, of A. Third, we choose a subset J of $[n]-A$. Fourth, we choose all functions f such that f is surjective from $[n]$ to A, from A to K and from J to $A-K$ when the domain of f is restricted to $[n]$, A and J respectively. Then the number of f is

$$\sum_{a=1}^{n}\binom{n}{a}\binom{a}{k}\left\langle {a \atop k}\right\rangle \sum_{j=0}^{n-a}\binom{n-a}{j}\left\langle {j \atop a-k}\right\rangle k^{n-a-j}. \qquad \text{(M)}$$

Before computing the expectation of the image size of $f \circ f$, we introduce a very useful lemma.

Lemma 16. Let n and k be nonnegative integers,

Then $\left\langle {n \atop k}\right\rangle = (-1)^k \sum_{j=0}^{k}(-1)^j\binom{k}{j}j^n$ (Here, we put $0^0=1$.)

Proof. See p.19 in [2]. □

Let's compute the expectation of the image size of a iterated function $f \circ f$.

Lemma 17. Let $1 \le n$. If f is a random function from $[n] \to [n]$, then, the image size of the iterated function $f \circ f : [n] \to [n]$ is expected to be

$$\frac{\sum_{a=1}^{n}(-1)^a\binom{n}{a}\sum_{k=1}^{a}k\binom{a}{k}\sum_{p=0}^{k}(-1)^p\binom{k}{p}p^a\sum_{q=0}^{a-k}(-1)^q\binom{a-k}{q}(q+k)^{n-a}}{n^n}.$$

Proof. $\dfrac{\sum_{f:[n]\to[n]} |\mathrm{Im}(f\circ f)|}{n^n}$

by (M),

$$= \frac{\sum_{k=1}^{n} k \sum_{a=1}^{n} \binom{n}{a}\binom{a}{k}\left\langle {a \atop k} \right\rangle \sum_{j=0}^{n-a} \binom{n-a}{j}\left\langle {j \atop a-k} \right\rangle k^{n-a-j}}{n^n}$$

by lemma 16,

$$= \frac{\sum_{k=1}^{n} k \sum_{a=1}^{n} \binom{n}{a}\binom{a}{k}\left\langle {a \atop k} \right\rangle \sum_{j=0}^{n-a} \binom{n-a}{j}\left[(-1)^{a-k}\sum_{q=0}^{a-k}(-1)^q\binom{a-k}{q}q^j\right] k^{n-a-j}}{n^n}$$

$$= \frac{\sum_{k=1}^{n} k \sum_{a=1}^{n} \binom{n}{a}\binom{a}{k}\left\langle {a \atop k} \right\rangle (-1)^{a-k}\sum_{q=0}^{a-k}(-1)^q\binom{a-k}{q}\sum_{j=0}^{n-a} \binom{n-a}{j} q^j k^{n-a-j}}{n^n}$$

$$= \frac{\sum_{k=1}^{n} k \sum_{a=1}^{n} \binom{n}{a}\binom{a}{k}\left\langle {a \atop k} \right\rangle (-1)^{a-k}\sum_{q=0}^{a-k}(-1)^q\binom{a-k}{q}(q+k)^{n-a}}{n^n}$$

by lemma 16,

$$= \frac{\sum_{k=1}^{n} k \sum_{a=1}^{n} \binom{n}{a}\binom{a}{k}\left[(-1)^{k}\sum_{p=0}^{k}(-1)^p\binom{k}{p}p^a\right](-1)^{a-k}\sum_{q=0}^{a-k}(-1)^q\binom{a-k}{q}(q+k)^{n-a}}{n^n}$$

$$= \frac{\sum_{a=1}^{n} \binom{n}{a} \sum_{k=1}^{a} k\binom{a}{k}\left[(-1)^{k}\sum_{p=0}^{k}(-1)^p\binom{k}{p}p^a\right](-1)^{a-k}\sum_{q=0}^{a-k}(-1)^q\binom{a-k}{q}(q+k)^{n-a}}{n^n}$$

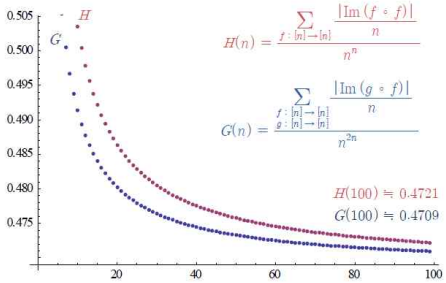
$$= \frac{\sum_{a=1}^{n} (-1)^a\binom{n}{a} \sum_{k=1}^{a} k\binom{a}{k}\sum_{p=0}^{k}(-1)^p\binom{k}{p}p^a\sum_{q=0}^{a-k}(-1)^q\binom{a-k}{q}(q+k)^{n-a}}{n^n}. \qquad \square$$

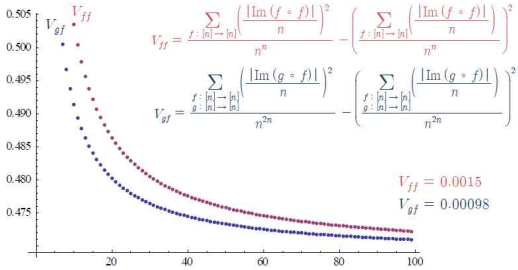
By lemma 17, the proportional image size of a double iterated random function $f\circ f : [n] \to [n]$ is expected to be

$$\frac{\sum_{f:[n]\to[n]} \frac{|\mathrm{Im}(f\circ f)|}{n}}{n^n} = \frac{\sum_{a=1}^{n} (-1)^a\binom{n}{a} \sum_{k=1}^{a} \frac{k}{n}\binom{a}{k}\sum_{p=0}^{k}(-1)^p\binom{k}{p}p^a\sum_{q=0}^{a-k}(-1)^q\binom{a-k}{q}(q+k)^{n-a}}{n^n}. \text{(N)}$$

Despite much effort, we could not determine the limit of (N) as $n \to \infty$. But we conjecture that the limit is $1-\left(\frac{1}{e}\right)^{1-\left(\frac{1}{e}\right)}$ through our experience and computing by Mathematica. Also, we conjecture that the limit of the variance is zero.

Conjecture 18. If f is a random function from $[n] \to [n]$, then, the image size of $f \circ f : [n] \to [n]$ is expected to be $n\left(1-\left(\frac{1}{e}\right)^{1-\left(\frac{1}{e}\right)}\right) \fallingdotseq (0.4685)n$ if n is sufficiently large.

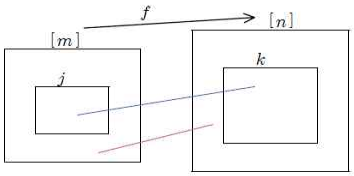




By Mathematica

4. Expectation of inverse image size of random function

Given a subset $[k]$ of $[n]$, the number of functions $f : [m] \to [n]$ such that the inverse image size of $[k]$ by f is j is $\binom{m}{j} k^j (n-k)^{m-j}$. (O)



Theorem 19. Let m, n and k be positive integers, If $k \le n$, then the inverse image size of $[k]$ by a random function $f : [m] \to [n]$ is expected to be $\frac{m}{n}k$.

Proof. If $k \ne n$, then, by lemma 7 and (O),

$$\frac{\sum_{f:[m]\to[n]} |f^{-1}([k])|}{n^m} = \frac{\sum_{j=0}^{m} j\binom{m}{j} k^j (n-k)^{m-j}}{n^m} = \frac{km\,n^{m-1}}{n^m} = \frac{km}{n}. \qquad \square$$

Example 20. Consider all functions $(a\ b\ c)$ from $[3]$ to $[4]$, where $(a\ b\ c)$ means $1 \mapsto a$, $2 \mapsto b$, $3 \mapsto c$.

$(a\ b\ c)$	$\lvert(a\ b\ c)^{-1}[2]\rvert$						
(1 1 1)	3	(2 1 1)	3	(3 1 1)	2	(4 1 1)	2
(1 1 2)	3	(2 1 2)	3	(3 1 2)	2	(4 1 2)	2
(1 1 3)	2	(2 1 3)	2	(3 1 3)	1	(4 1 3)	1
(1 1 4)	2	(2 1 4)	2	(3 1 4)	1	(4 1 4)	1
(1 2 1)	3	(2 2 1)	3	(3 2 1)	2	(4 2 1)	2
(1 2 2)	3	(2 2 2)	3	(3 2 2)	2	(4 2 2)	2
(1 2 3)	2	(2 2 3)	2	(3 2 3)	1	(4 2 3)	1
(1 2 4)	2	(2 2 4)	2	(3 2 4)	1	(4 2 4)	1
(1 3 1)	2	(2 3 1)	2	(3 3 1)	1	(4 3 1)	1
(1 3 2)	2	(2 3 2)	2	(3 3 2)	1	(4 3 2)	1
(1 3 3)	1	(2 3 3)	1	(3 3 3)	0	(4 3 3)	0
(1 3 4)	1	(2 3 4)	1	(3 3 4)	0	(4 3 4)	0
(1 4 1)	2	(2 4 1)	2	(3 4 1)	1	(4 4 1)	1
(1 4 2)	2	(2 4 2)	2	(3 4 2)	1	(4 4 2)	1
(1 4 3)	1	(2 4 3)	1	(3 4 3)	0	(4 4 3)	0
(1 4 4)	1	(2 4 4)	1	(3 4 4)	0	(4 4 4)	0
Sum	32		32		16		16

We see that $\frac{\sum_{f:[3]\to[4]} |f^{-1}([2])|}{4^3} = \frac{32+32+16+16}{64} = \frac{96}{64} = \frac{3}{2}$. Also, by Theorem 19, $\frac{\sum_{f:[3]\to[4]} |f^{-1}([2])|}{4^3} = \frac{3}{4} \cdot 2 = \frac{3}{2}$. $\square$

5. Conclusion

(1), (2) and (3) hold for sufficiently large n.

(1) If $f : [n] \to [n]$ is a random function, then the image size of f is expected to be $\alpha_1 = n\left(1 - \frac{1}{e}\right)$.

(2) If f and g are random functions from $[n]$ to $[n]$, then, the image size of $g \circ f$ is expected to be $\alpha_2 = n\left(1 - \left(\frac{1}{e}\right)^{1-\left(\frac{1}{e}\right)}\right)$.

(3) If f is a random function from $[n]$ to $[n]$, then, we <u>conjecture</u> the image size of $f \circ f$ is expected to be $\alpha_2 = n\left(1 - \left(\frac{1}{e}\right)^{1-\left(\frac{1}{e}\right)}\right)$.

$$\frac{\alpha_2}{n} = 1 - \left(\frac{1}{e}\right)^{\frac{\alpha_1}{n}}.$$

(4) The inverse image size of $[k]$ by a random function $f : [m] \to [n]$ is expected to be $\frac{m}{n}k$ for $k \leq n$.

References

[1] 김홍종, 미적분학 1, 서울대학교출판부(2008).

[2] Herbert S. Wilf, generatingfunctionology, Academic Press, Inc.

[3] J. Hong, S. Moon, A Comparison of Cryptanalytic Tradeoff Algorithms.

국문초록.

m이 양의 정수일 때, 우리는 집합 $\{1, 2, ..., m\}$을 $[m]$으로 나타낸다. 또한 함수 $f : [m] \to [n]$의 이미지 사이즈는 함수 f의 치역의 원소의 개수를 의미하고 $|f([m])|$ 또는 $|\mathrm{Im}\, f|$으로 나타내자. 집합 $[m]$에서 집합 $[n]$으로 가는 모든 함수의 모임을 생각하자. 각 함수가 선택될 확률이 동일할 때, 집합 $[m]$에서 집합 $[n]$으로 가는 랜덤함수는 우리가 이 집합에서 선택할 함수이다. 여기서 랜덤함수는 이미 선택된 함수가 아니라는 것에 주의해야 한다. 이 논문에서 우리는 두 번 합성된 랜덤함수의 이미지 사이즈의 기댓값을 계산해 본다.

주요 어휘 : 랜덤함수, 이미지 사이즈의 기댓값

학번 : 2002-23256