短いループのできる確率

後藤達哉 (筑波大学理工学群数学類3年)

2018年12月8日

数物セミナー愛媛談話会

https://fujidig.github.io/201812-short-period/201812-short-period.pdf

この発表で話すこと

- $\underline{m} = \{0, 1, 2, \ldots, m-1\}$とおく
- $\underline{m}$から$\underline{m}$への写像fが無作為に与えられたときにfによる反復列の周期に関する期待値や確率を議論する
- そのようなfはm^m通りあるのでそのそれぞれが等確率であるとする
- mが十分大きいときに興味があるので$m \to \infty$の極限を考える

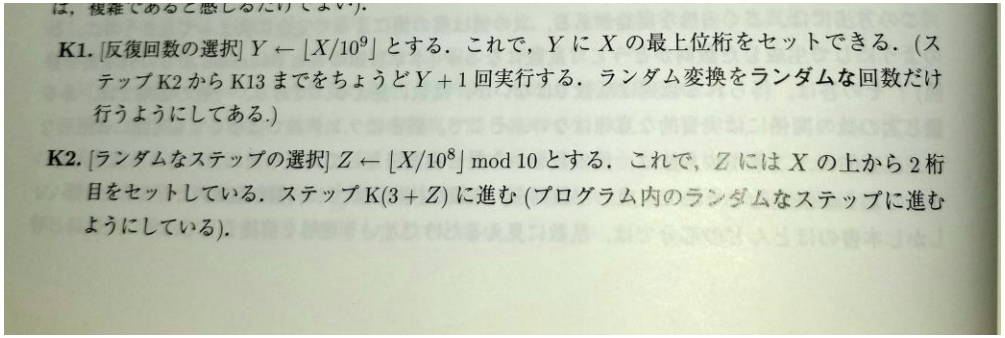
このテーマのモチベーション

- このテーマは疑似乱数の発生と関係してくる
- 一般的にコンピュータは(外部からノイズを入力するということをしない限り)乱数を発生させることはできない
- そこで確定的な計算によって乱数っぽいもの(疑似乱数)を作っている
- 疑似乱数は固定された写像fを使った反復列によって定められる

このテーマのモチベーション

- 疑似乱数の生成に使われるfは「でたらめな写像」を持ってくれば良いだろうという誤解がコンピュータの最初期にはあった
- しかし，「でたらめな写像」は短い周期を持ちやすいので，疑似乱数にふさわしくない
- その事実を定量的に述べる一つの議論がこの講演である

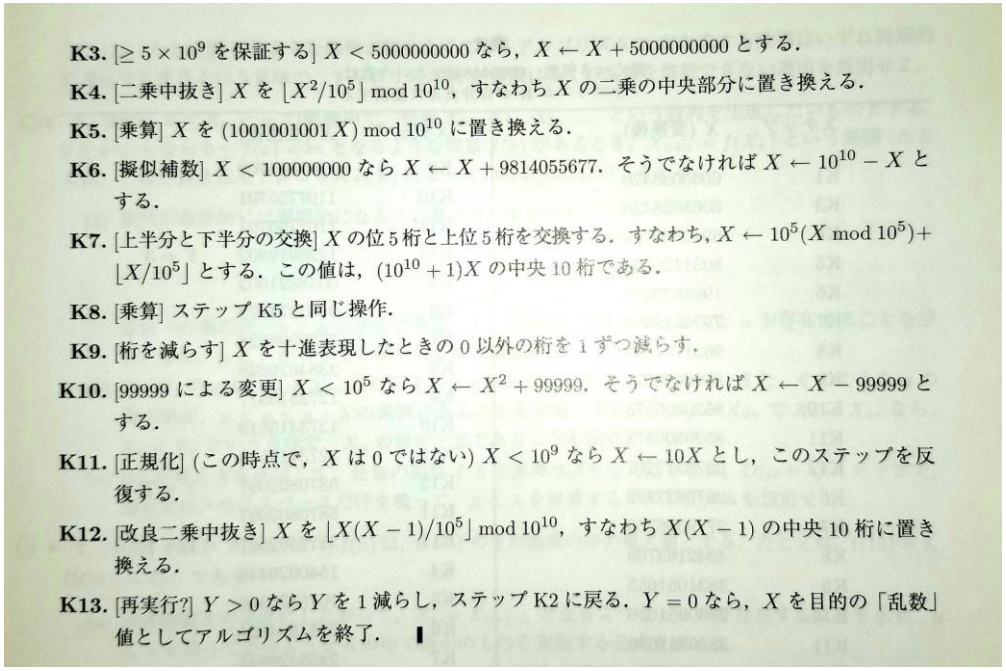
Knuthが考えたでたらめな写像



は，複雑であると感じるだけでよい）．

K1. [反復回数の選択] $Y \leftarrow \lfloor X/10^9 \rfloor$ とする．これで，Y に X の最上位桁をセットできる．(ステップ K2 から K13 までをちょうど $Y+1$ 回実行する．ランダム変換をランダムな回数だけ行うようにしてある．)

K2. [ランダムなステップの選択] $Z \leftarrow \lfloor X/10^8 \rfloor \bmod 10$ とする．これで，Z には X の上から 2 桁目をセットしている．ステップ K$(3+Z)$ に進む (プログラム内のランダムなステップに進むようにしている)．



K3. [$\geq 5 \times 10^9$ を保証する] $X < 5000000000$ なら，$X \leftarrow X + 5000000000$ とする．

K4. [二乗中抜き] X を $\lfloor X^2/10^5 \rfloor \bmod 10^{10}$，すなわち X の二乗の中央部分に置き換える．

K5. [乗算] X を $(1001001001\,X) \bmod 10^{10}$ に置き換える．

K6. [擬似補数] $X < 100000000$ なら $X \leftarrow X + 9814055677$．そうでなければ $X \leftarrow 10^{10} - X$ とする．

K7. [上半分と下半分の交換] X の位 5 桁と上位 5 桁を交換する．すなわち，$X \leftarrow 10^5(X \bmod 10^5) + \lfloor X/10^5 \rfloor$ とする．この値は，$(10^{10}+1)X$ の中央 10 桁である．

K8. [乗算] ステップ K5 と同じ操作．

K9. [桁を減らす] X を十進表現したときの 0 以外の桁を 1 ずつ減らす．

K10. [99999 による変更] $X < 10^5$ なら $X \leftarrow X^2 + 99999$．そうでなければ $X \leftarrow X - 99999$ とする．

K11. [正規化] (この時点で，X は 0 ではない) $X < 10^9$ なら $X \leftarrow 10X$ とし，このステップを反復する．

K12. [改良二乗中抜き] X を $\lfloor X(X-1)/10^5 \rfloor \bmod 10^{10}$，すなわち $X(X-1)$ の中央 10 桁に置き換える．

K13. [再実行?] $Y > 0$ なら Y を 1 減らし，ステップ K2 に戻る．$Y = 0$ なら，X を目的の「乱数」値としてアルゴリズムを終了． ∎

[2]の3.1節より引用.
このアルゴリズムはKnuthが1959年に考えた.

このアルゴリズムで得られる写像を使って列を作ると初期値によっては周期が1や3になりえる．つまり疑似乱数としては使い物にならない

(Knuthでさえ，でたらめな写像が疑似乱数に使えると誤解した)

問題1

問題1

写像$f:\underline{m} \to \underline{m}$と$x_0 \in \underline{m}$が無作為に与えられるときに列$x_n = f^n(x_0)$の周期が1になる確率を求めよ.

問題1

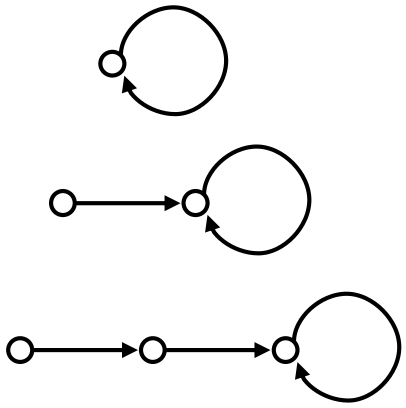
写像$f:\underline{m}\to\underline{m}$と$x_0\in\underline{m}$が無作為に与えられるときに列$x_n=f^n(x_0)$の周期が1になる確率を求めよ.

$x_{i+k}=x_i$となる$i,k\geq 0$がある.
そのようなiの最小値をμ, kの最小値をλとする.
λが列の周期である.

$$P(\mu=0,\lambda=1)=\frac{1}{m}$$

$$P(\mu=1,\lambda=1)=\left(1-\frac{1}{m}\right)\frac{1}{m}$$

$$P(\mu=2,\lambda=1)=\left(1-\frac{1}{m}\right)\left(1-\frac{2}{m}\right)\frac{1}{m}$$



問題1

写像$f:\underline{m}\to\underline{m}$と$x_0\in\underline{m}$が無作為に与えられるときに列$x_n=f^n(x_0)$の周期が1になる確率を求めよ.

一般に

$$P(\mu=\mu_0,\lambda=1)=\left(1-\frac{1}{m}\right)\left(1-\frac{2}{m}\right)\ldots\left(1-\frac{\mu_0}{m}\right)\frac{1}{m}$$

よって

$$P(\lambda=1)=\frac{1}{m}\sum_{i\geq 0}\left(1-\frac{1}{m}\right)\left(1-\frac{2}{m}\right)\ldots\left(1-\frac{i}{m}\right)$$

□ を$Q(m)$とおく.

問題1

写像$f:\underline{m} \to \underline{m}$と$x_0 \in \underline{m}$が無作為に与えられるときに列$x_n = f^n(x_0)$の周期が1になる確率を求めよ.

$$Q(m) = \sum_{i \geq 0} \left(1 - \frac{1}{m}\right)\left(1 - \frac{2}{m}\right) \ldots \left(1 - \frac{i}{m}\right)$$

はRamanujanのQ functionと呼ばれる.

Fact

$$Q(m) = \sqrt{\frac{\pi m}{2}} - \frac{1}{3} + O(m^{-\frac{1}{2}})$$

問題1

Factより,

$$P(\lambda = 1) \sim \sqrt{\frac{\pi}{2m}} \sim \frac{1.25}{\sqrt{m}}$$

$\rightsquigarrow$ mが十分大きいときには周期が1になる確率はほぼ0.

問題2

問題2

写像$f:\underline{m} \to \underline{m}$と$x_0 \in \underline{m}$が無作為に与えられるときに列$x_n = f^n(x_0)$の周期の期待値を求めよ.

問題2

写像$f:\underline{m}\to\underline{m}$と$x_0\in\underline{m}$が無作為に与えられるときに列$x_n=f^n(x_0)$の周期の期待値を求めよ.

一般に

$$P(\mu=\mu_0,\lambda=\lambda_0)=\frac{1}{m}\prod_{1\le k<\mu_0+\lambda_0}\left(1-\frac{k}{m}\right).$$

よって

$$E[\lambda]=\sum_{\substack{1\le\lambda_0\\0\le\mu_0}}\frac{\lambda_0}{m}\prod_{1\le k<\mu_0+\lambda_0}\left(1-\frac{k}{m}\right).$$

問題2

写像$f:\underline{m}\to\underline{m}$と$x_0\in\underline{m}$が無作為に与えられるときに列$x_n=f^n(x_0)$の周期の期待値を求めよ.

$$E[\lambda]=\sum_{\substack{1\le\lambda_0\\0\le\mu_0}}\frac{\lambda_0}{m}\prod_{1\le k<\mu_0+\lambda_0}\left(1-\frac{k}{m}\right)$$

$$=\frac{1}{m}\left\{1+(1+2)\left(1-\frac{1}{m}\right)+(1+2+3)\left(1-\frac{1}{m}\right)\left(1-\frac{2}{m}\right)+\cdots\right\}$$

$$=\frac{1}{m}\sum_{n\ge0}\frac{(n+1)(n+2)}{2}\prod_{1\le k<n}\left(1-\frac{k}{m}\right)$$

問題2

写像$f:\underline{m}\to\underline{m}$と$x_0\in\underline{m}$が無作為に与えられるときに列$x_n=f^n(x_0)$の周期の期待値を求めよ.

Lemma

$$f(a_0,a_1,\dots)=\sum_{n\geq 0}a_n\prod_{k=1}^{n}(1-\frac{k}{m})$$

なら

$$f(a_0,a_1,\dots)=a_0+f(a_1,a_2,\dots)-f(a_1,2a_2,\dots)\,/\,m$$

Lemmaに$a_n=\frac{n+1}{2}$を代入すると

$$\mathrm{E}[\lambda]=\frac{1}{m}\sum_{n\geq 0}\frac{(n+1)(n+2)}{2}\prod_{1\leq k<n}\left(1-\frac{k}{m}\right)=\frac{1+Q(m)}{2}.$$

問題2

写像$f:\underline{m}\to\underline{m}$と$x_0\in\underline{m}$が無作為に与えられるときに列$x_n=f^n(x_0)$の周期の期待値を求めよ.

よって

$$\mathrm{E}[\lambda]=\frac{1+Q(m)}{2}\sim\sqrt{\frac{\pi m}{8}}+\frac{1}{3}.$$

つまり周期λの期待値は$\sqrt{m}$に比例する程度

問題3

問題3

写像$f:\underline{m} \to \underline{m}$が無作為に与えられるとき，$f$が不動点を持つ確率を求めよ．

問題3

写像$f:\underline{m} \to \underline{m}$が無作為に与えられるとき，$f$が不動点を持つ確率を求めよ．

不動点とは$f(x) = x$となる$x \in \underline{m}$のこと．
すべてのxで$f(x) \neq x$となるfの個数は$(m-1)^m$
一方，すべてのfの個数はm^m．
よって

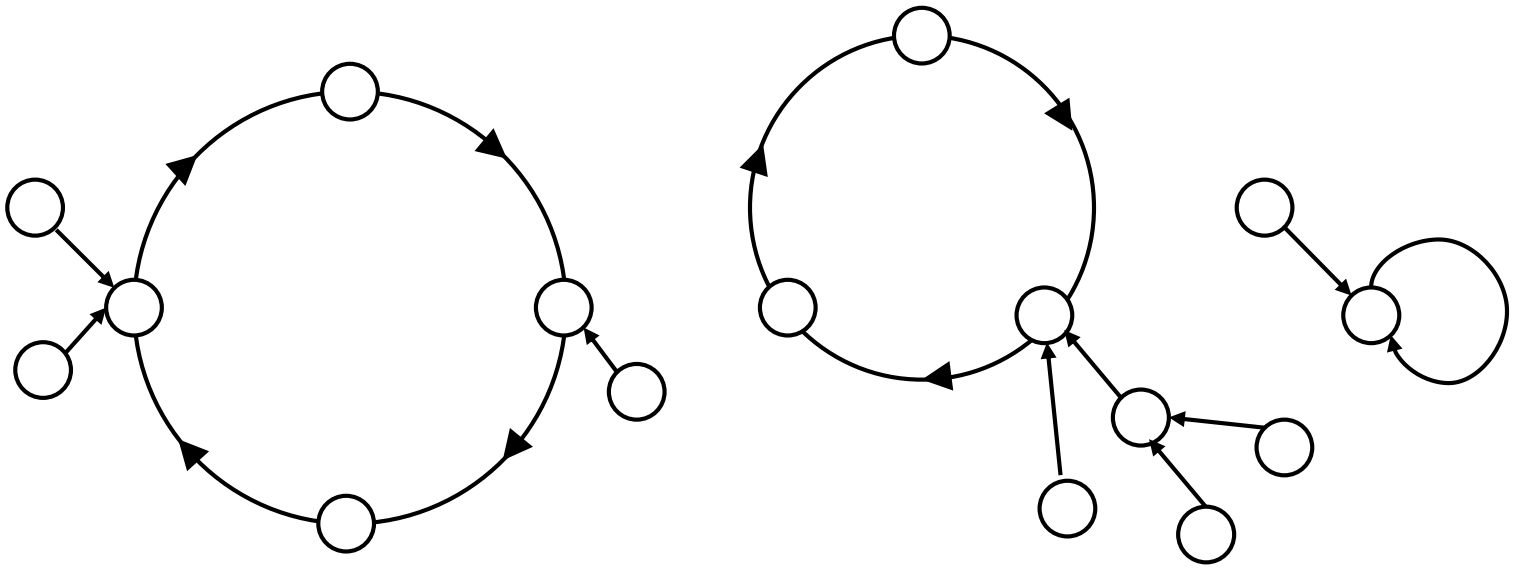
$$P(\text{不動点を持つ}) = 1 - \frac{(m-1)^m}{m^m}$$
$$= 1 - \left(1 - \frac{1}{m}\right)^m$$

$m \to \infty$の極限は$1 - e^{-1} = 0.632\cdots$である．

問題4

…の前に観察

- $0, 1, 2, \ldots, m-1$を頂点として$f(i) = j$ならiからjへの有向辺を伸ばしたグラフを考える
- このグラフはいくつかのサイクルに木を生やしたものになっている

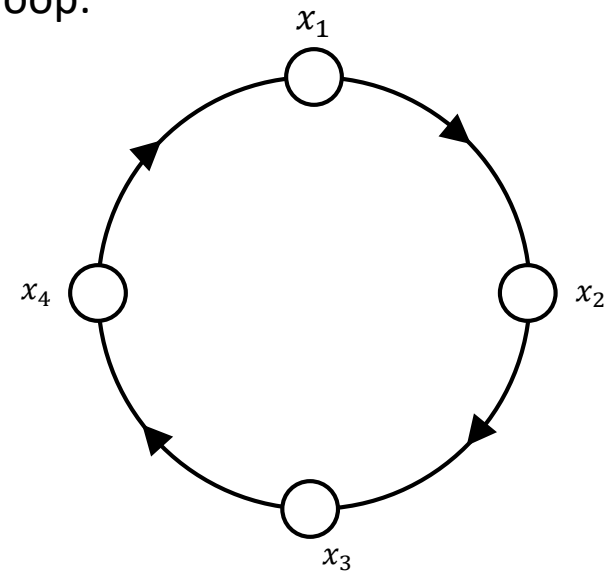


問題4

問題3を一般化しよう.

fのl-loopとは$\underline{m}$の相異なる元$x_1, x_2, \ldots, x_l$であって$f(x_1) = x_2, \ldots, f(x_l) = x_1$をみたすもののこと.

4-loop:



写像$f: \underline{m} \to \underline{m}$が無作為に与えられるとき，$n$以下のloopができる確率を求めよ.

問題4

写像$f:\underline{m} \to \underline{m}$が無作為に与えられるとき，$n$以下のloopができる確率を求めよ．

かんたんのため，$n = 2$で解く．

Fact (包除原理)

Iを有限全順序集合，各$i \in I$に対しA_iを事象とする．このとき

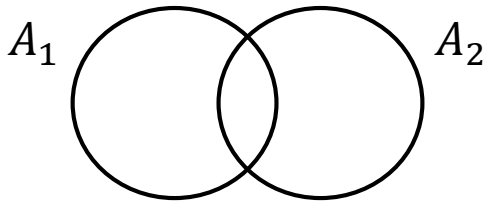
$$P\left(\bigcup_{i\in I} A_i\right) = \sum_{k=1}^{|I|} (-1)^{k-1} \sum_{i_1<\cdots<i_k} P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k})$$

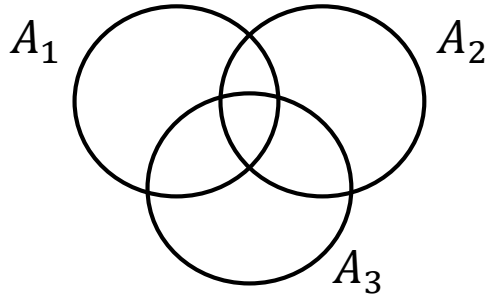
問題4

写像$f: \underline{m} \to \underline{m}$が無作為に与えられるとき，$n$以下のloopができる確率を求めよ．

包除原理は$|I| = 2, 3$のときは次の式を表す:

- $P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$
- $P(A_1 \cup A_2 \cup A_3) = P(A_1) + P(A_2) + P(A_3) - P(A_1 \cap A_2) - P(A_1 \cap A_3) - P(A_2 \cap A_3) + P(A_1 \cap A_2 \cap A_3)$





問題4

写像$f:\underline{m}\to\underline{m}$が無作為に与えられるとき，$n$以下のloopができる確率を求めよ.

- $\left[\underline{m}\right]^2=\{\{x,y\}\subset\underline{m}\,|\,x\neq y\}$ とおく
- $x\in\underline{m}$に対し$A_x=\{f:\underline{m}\to\underline{m}\,|\,x\text{は}f\text{の不動点}\}$
- $\{x,y\}\in\left[\underline{m}\right]^2$に対し

  $A_{\{x,y\}}=\{f:\underline{m}\to\underline{m}\,|\,\{x,y\}\text{は}f\text{の}2-\text{loop}\}$

  と定める.

問題4

写像$f:\underline{m}\to\underline{m}$が無作為に与えられるとき，$n$以下のloopができる確率を求めよ．

すると

$$P(\exists 1-\text{loop}\vee\exists 2-\text{loop})$$

$$=P\left(\bigcup_{t\in\underline{m}\cup[\underline{m}]^2}A_t\right)$$

$$=\sum_{k=1}^{|I|}(-1)^{k-1}\sum_{\substack{t_1<\cdots<t_k\\ \in\underline{m}\cup[\underline{m}]^2}}P(A_{t_1}\cap A_{t_2}\cap\cdots A_{t_k})$$

$$=\sum_{k=1}^{}(-1)^{k-1}\sum_{\substack{r+s=k\\ r+2s\le m}}\sum_{\substack{x_1<\cdots<x_r\in\underline{m}\\ p_1<\cdots<p_s\in[\underline{m}]^2}}P(A_{x_1}\cap\cdots\cap A_{x_r}\cap A_{p_1}\cap\cdots\cap A_{p_s})$$

① ② ③

問題4

写像$f:\underline{m} \to \underline{m}$が無作為に与えられるとき，$n$以下のloopができる確率を求めよ．

①について計算．固定された$x_1, \dots, x_r, p_1, \dots, p_s$について

$$\sum_{\substack{x_1<\cdots<x_r\in\underline{m} \\ p_1<\cdots<p_s\in[\underline{m}]^2}} P(A_{x_1} \cap \cdots \cap A_{x_r} \cap A_{p_1} \cap \cdots \cap A_{p_s})$$

$$= \frac{m^{m-r-2s}}{m^m} = \frac{1}{m^{r+2s}}$$

問題4

写像$f:\underline{m}\to\underline{m}$が無作為に与えられるとき，$n$以下のloopができる確率を求めよ．

②で走らせているものの個数を数える．

$$\#\left\{\left((x_1,\dots,x_r),(p_1,\dots,p_s)\right)\in\underline{m}^r\times\left(\left[\underline{m}\right]^2\right)^s \;\middle|\; \begin{matrix} x_1<\cdots<x_r, p_1<\cdots<p_s, \\ \{x_1\},\dots,\{x_r\},p_1,\dots,p_s\text{はpairwise disjoint}\end{matrix}\right\}$$

$$=C(m,r)\frac{P(m-r,2s)}{(2!)^s s!}$$

ここに$C(-,-)$は組み合わせの数，$P(-,-)$は順列の数を表す記号．

よって，①と②とから③が計算される．

$$\sum_{\substack{x_1<\cdots<x_r\in\underline{m}\\ p_1<\cdots<p_s\in[\underline{m}]^2}} P\left(A_{x_1}\cap\cdots\cap A_{x_r}\cap A_{p_1}\cap\cdots\cap A_{p_s}\right)=C(m,r)\frac{P(m-r,2s)}{2^s s!}\frac{1}{m^{r+2s}}$$

$$=\frac{1}{2^s r!s!}\left(1-\frac{1}{m}\right)\left(1-\frac{2}{m}\right)\dots\left(1-\frac{r+2s-1}{m}\right)$$

問題4

写像$f:\underline{m}\to\underline{m}$が無作為に与えられるとき，$n$以下のloopができる確率を求めよ．

よって

$$P(\exists 1-\text{loop}\vee\exists 2-\text{loop})$$

$$=\sum_{k=1}(-1)^{k-1}\sum_{\substack{r+s=k\\ r+2s\le m}}(③)$$

$$=\sum_{k=1}(-1)^{k-1}\sum_{\substack{r+s=k\\ r+2s\le m}}\frac{1}{2^s r!s!}\left(1-\frac{1}{m}\right)\left(1-\frac{2}{m}\right)\dots\left(1-\frac{r+2s-1}{m}\right)$$

$$=\sum_{\substack{r+s\ge 1\\ r+2s\le m}}\frac{(-1)^{r+s-1}}{2^s r!s!}\left(1-\frac{1}{m}\right)\left(1-\frac{2}{m}\right)\dots\left(1-\frac{r+2s-1}{m}\right)$$

問題4

写像$f:\underline{m} \to \underline{m}$が無作為に与えられるとき，$n$以下のloopができる確率を求めよ．

$$P(\exists 1-\text{loop} \vee \exists 2-\text{loop})$$

$$= \sum_{\substack{r+s\geq 1 \\ r+2s\leq m}} \frac{(-1)^{r+s-1}}{2^s r!s!}\left(1-\frac{1}{m}\right)\left(1-\frac{2}{m}\right)\dots\left(1-\frac{r+2s-1}{m}\right)$$

$m \to \infty$とすると

$$\sum_{r+s\geq 1} \frac{(-1)^{r+s-1}}{2^s r!\, s!} = 1 - \sum_{r\geq 0}\frac{(-1)^r}{r!}\sum_{s\geq 0}\frac{(-1)^s}{2^s s!}$$

$$= 1 - \mathrm{e}^{-1}\mathrm{e}^{-\frac{1}{2}}$$

$$= 1 - \mathrm{e}^{-\frac{3}{2}}$$

(ルベーグの収束定理とフビニの定理を使った)

問題4

写像$f:\underline{m}\to\underline{m}$が無作為に与えられるとき，$n$以下のloopができる確率を求めよ．

一般のnでは

$$P_n := \lim_{m\to\infty} P\left(\left\{f:\underline{m}\to\underline{m}\middle| f\text{が}n\text{以下のloopを持つ}\right\}\right)$$

$$= 1-e^{-H_n}\quad\left(H_n = 1+\frac{1}{2}+\cdots+\frac{1}{n}\right)$$

となる．

$$P_1 = 0.632\cdots$$
$$P_2 = 0.776\cdots$$
$$P_3 = 0.840\cdots$$
$$P_{10} = 0.946\cdots$$
$$P_{100} = 0.994\cdots$$

問題5

問題5

写像$f:\underline{m} \to \underline{m}$が無作為に与えられるとき，$f$の一番短いloopの長さの期待値は？

（未解決？）

具体例

- $m = 2^{24} = 16777216$としてfは次のものを採用
- $f(x) = \left\lfloor \frac{(ax+b) \bmod 2^{64}}{2^{40}} \right\rfloor$

where $a = 1020887724600906955 1, b = 17024059796969606487$

- このときループの長さは

$$4051, 2523, 259, 197, 165, 140, 139, 2$$

となった (これで全て)

なお, $\sqrt{\frac{\pi m}{8}} = 2566.\cdots$

参考文献

[1] Donald E. Knuth “The Art of Computer Programming, Volume 1: Fundamental Algorithms” Addison-Wesley Professional, 1997.

[2] Donald E. Knuth “The Art of Computer Programming, Volume 2: Seminumerical Algorithms” Addison-Wesley Professional, 2014.

本発表は主に[2]の3.1節の演習問題に依っている.

まとめ

- (問題1) 写像と初期値を無作為に選んだとき，周期が1になる確率は約 $\sqrt{\frac{\pi}{2m}}$
- (問題2) 周期の期待値は約 $\sqrt{\frac{\pi m}{8}}$
- (問題3) 写像を無作為に選んだとき不動点ができる確率は約 $1 - e^{-1} = 0.632\cdots$
- (問題4) n以下のloopができる確率は約 $1 - e^{-H_n}$ $(H_n = 1 + \frac{1}{2} + \cdots + \frac{1}{n})$
- (問題5) 一番短いloopの長さの期待値は？

- “The random numbers should not be generated with a method chosen at random.” (Knuth [2] 3.1節)

- ご清聴ありがとうございました