ポラードの ρ 素因数分解法

ρ 法はポラード（Pollard）によって発明された素因数分解法であり，十数桁程度の大きさまでの素因数を，確率的に探す乱択アルゴリズムである。

いま，合成数 N が与えられたとし，その素因数の 1 つを p とする。

$f(x)$ を多項式で表された関数とする。このとき， a を $0 \leqq a < p$ を満たす整数とし

$$x_0 = a, \quad x_{n+1} = \mathrm{mod}(f(x_n), N)$$

で定義される数列 $\{x_n\}$ を考える。ただし， $\mathrm{mod}(x, N)$ は x を N で割った剰余を表すものとする。

さらに， $y_n = \mathrm{mod}(x_n, p)$ とする。

ここで， $0 \leqq y_n < p$ が成り立つから，数列 $\{y_n\}$ から連続した $p+1$ 個の項を取り出せば，鳩ノ巣原理により，少なくとも 1 組の等しい項が存在する。すなわち

$$y_J = y_I, \quad 1 \leqq I < J \leqq p+1$$

となる I, J が存在する。 $J - I = t$ （$t \leqq p$）とおくと，

$$y_{I+t} = y_I$$

漸化式により y_n から y_{n+1} が一意的に決定されるから，$n \geqq I$ では，t は数列 $\{y_n\}$ の周期となる。（ただし，y_n から y_{n-1} は一意的に決定されるわけではないことに注意したい。）

数列 $\{y_n\}$ において， $n \geqq I$ における正で最小の周期を T で表し，これを単に周期と呼ぶことにする。

このとき， $i \geqq I$ である任意の i について

$$y_{i+T} - y_i = 0$$

したがって，

$$\mathrm{mod}(x_{i+T} - x_i, p) = 0$$

これより， $x_{i+T} - x_i$ は p の倍数であるから， $x_{i+T} - x_i$ と N の最大公約数は p の倍数である。このとき，$1, N$ 以外の最大公約数が見つかれば，それは， N の自明でない約数となる。

ここで， $0 < (i+T) - i = T \leqq p$ であるから

$$\mathrm{mod}(x_j - x_i, p) = 0, \quad I \leqq i < j \leqq I + p$$

となる i, j が存在することになる。このとき， $0 < j - i \leqq p$ である。

いま， $\mathrm{mod}(x_j - x_i, p) = 0$ となる i, j が見つかったとしよう。このとき

$$\mathrm{mod}(x_j, p) = \mathrm{mod}(x_i, p)$$

$f(x)$ は x の多項式であるから，

$$\mathrm{mod}(f(x_j), p) = \mathrm{mod}(f(x_i), p)$$

よって $\mathrm{mod}(f(x_j) - f(x_i), p) = 0$

すなわち

$$\mathrm{mod}(x_{j+1} - x_{i+1}, p) = 0$$

これより

$$\mathrm{mod}(x_j - x_i,\, p) = 0 \;\Rightarrow\; \mathrm{mod}(x_{j+1} - x_{i+1},\, p) = 0 \quad \cdots①$$

が成り立つ。
$I \leqq i < j$ である場合を考える。このとき，数列$\{y_n\}$は循環の輪の中にあるから，①を繰り返し用いると

$$\mathrm{mod}(x_{j+1} - x_{i+1},\, p) = 0 \;\Rightarrow\; \mathrm{mod}(x_j - x_i,\, p) = 0$$

が得られる。よって

$$\mathrm{mod}(x_j - x_i,\, p) = 0 \;\Leftrightarrow\; \mathrm{mod}(x_{j+1} - x_{i+1},\, p) = 0 \quad \cdots②$$

が成り立つ。

循環の輪の中では，$\mathrm{mod}(x_j - x_i,\, p) = 0$ となる $i,\, j$ の組を見つけるのに，$I \leqq i < j \leqq I + T - 1$ を満たす

全ての組について$|x_i - x_j|$と N の最大公約数を求める必要はなく，②により

$$j - i = 1, 2, 3, \cdots\cdots, \left\lfloor \frac{T}{2} \right\rfloor$$

となる $i,\, j$ の組のそれぞれ 1 ずつだけ$|x_i - x_j|$と N の最大公約数を求めればよい。

したがって，$I \leqq i < j$ ならば，i を I に固定するのが最も効率的である。しかし，I を知ることができない。よって，ポラードは，$|x_2 - x_1|$，$|x_4 - x_2|$，$|x_6 - x_3|$，……と N との最大公約数を求めた。
ブレントは，このアルゴリズムを改良して，項数が 2 倍ごとに x_i を固定値とするアルゴリズムを考えた。
さらに，ブレントのアルゴリズムを独自に改良することを考えた。
ポラードの ρ 素因数分解法は，次にある上下に並んだ添字の項の差を用いる。下の行が空白の場合は用いない。添字の差は，1，2，3，……と連続する。

ポラードの元のアルゴリズムは，次の添字に従う。一度計算した項をすべて記憶することができないから，上の列と下の列を添字に持つ乱数列を同時に 2 本生成する。

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
	1		2		3		4		5		6		7		8		9		10		11		12		13		14		15	

ブレントによる改良アルゴリズムは，次の添字に従う。既に生成した項を 1 項だけ記憶することにより，乱数列を 1 本だけ生成して済ます。

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
	1		2				4	4					8	8	8	8															16	16	16	16	16	16	16	16

ブレントの方法に独自の改良を加えたアルゴリズムは，既に生成した項を何項か記憶することにより，若干の高速化を図ったもので
4 項だけ記憶する場合は

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
				4	4	4	4		5		6		7		8			10	10			12	12			14	14			16

8 項だけ記憶する場合は

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
								8	8	8	8	8	8	8	8		9		10		11		12		13		14		15	

ところで，この最大公約数が N になった場合は，このアルゴリズムは失敗する。なぜなら，このとき $x_{2n} = x_n$ となるため，それ以降の数列がそれまでの繰り返しになってしまうからである。

註　数列 $\{y_n\}$ は，最初は周期的でなくても途中から周期的になる（循環の輪に入る）様子が，ギリシア文字の ρ に似ていることから，ρ 法と呼ばれている。

以上の考察では，$\mathrm{mod}(x_j - x_i,\ p) = 0$ となる $i,\ j$ の組を見つけるのに，p のオーダーの回数の探索を行わなければならない。ところが，次の誕生日のパラドックスにより，$\sqrt{p}$ 程度のオーダーの回数の探索で見つけられる可能性がある。しかし，これは証明されたものではなく，実験結果から正しいのではないかと考えれれている。

誕生日のパラドックス

23 人の人間が居れば，その中に同じ誕生日の 2 人(以上)がいる確率は約 $\frac{1}{2}$ である。これを，誕生日のパラドックスという。一般に，n が十分大きいとき，n 種類の値をとる真の乱数が $\sqrt{n}$ 個あれば，乱数の中に同じ値を持つ 2 数(以上)が存在する確率は，$\frac{1}{2}$ に近い。

したがって，**数列 $\{y_n\}$ の項を $\sqrt{p}$ 個集めたものが，もしも真の乱数を $\sqrt{p}$ 個集めたものに近ければ，** $\sqrt{p}$ 個の項の中に，p を法として合同であるものが存在する確率は $\frac{1}{2}$ に近い。真の乱数については，その合同である 2 つの項の組を求めるのに，すべての 2 個の組み合わせである ${}_{\sqrt{p}}C_2 \approx \frac{p}{2}$ 通りを調べなければならない。しかし，$f(x)$ が多項式の場合は，性質①があるから，$\sqrt{p}$ 個程度の組み合わせを調べればよい。

誕生日のパラドックスの導出（誕生日攻撃）

真の乱数が n 種類の値を取るとき，k 個の乱数の中に同じ値を持つ 2 数(以上)が存在する確率を求めてみよう。

k 個の乱数がすべて異なる確率を p_1 とすると

$$p_1 = \frac{{}_nP_k}{n^k} = \frac{n(n-1)(n-2)\cdots\cdots(n-k+1)}{n^k}$$

$$=\left(1-\frac{0}{n}\right)\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\cdots\cdots\left(1-\frac{k-1}{n}\right)$$

よって

$$\log p_1=\log\left(1-\frac{0}{n}\right)+\log\left(1-\frac{1}{n}\right)+\log\left(1-\frac{2}{n}\right)+\cdots\cdots+\log\left(1-\frac{k-1}{n}\right)$$
$$=n\cdot\frac{1}{n}\left\{\log\left(1-\frac{0}{n}\right)+\log\left(1-\frac{1}{n}\right)+\log\left(1-\frac{2}{n}\right)+\cdots\cdots+\log\left(1-\frac{k-1}{n}\right)\right\}$$

n が十分大きく，k もある程度の大きいときを考えると

$$\log p_1\approx n\cdot\int_0^{\frac{k}{n}}\log(1-x)dx$$

x が十分に 0 に近いとき，$\log(1-x)\approx -x$ であるから，$\frac{k}{n}$ が十分に 0 に近いときを考えると

$$\log p_1\approx n\cdot\int_0^{\frac{k}{n}}(-x)dx=n\cdot\left[-\frac{1}{2}x^2\right]_0^{\frac{k}{n}}=-\frac{k^2}{2n}$$

よって

$$p_1\approx e^{-\frac{k^2}{2n}}$$

したがって，k 個の乱数の中に同じ値を持つ 2 数(以上)が存在する確率を p_2 とすると

$$p_2=1-p_1\approx 1-e^{-\frac{k^2}{2n}}$$

これより，n が十分大きい場合には

$k\approx\sqrt{n}$ のとき $p_2\approx 1-e^{-\frac{1}{2}}\approx 0.39$

$k\approx 2\sqrt{n}$ のとき $p_2\approx 1-e^{-2}\approx 0.86$

$k\approx 3\sqrt{n}$ のとき $p_2\approx 1-e^{-\frac{9}{2}}\approx 0.99$

$k\approx 4\sqrt{n}$ のとき $p_2\approx 1-e^{-8}\approx 0.9997$

となる。