数理解析研究所講究録
1127巻 2000年 115-128



# 定常離散無記憶情報源に対する乱数生成問題について

大濱靖匡
Yasutada Oohama
九州大学 大学院システム情報科学研究科
E-mail: oohama@csce.kyushu-u.ac.jp

***Abstract***- ある与えられた情報源からの出力系列（コイン乱数列）を変換して，所望の情報源からの出力系列（ターゲット乱数列）を生成あるいは近似することを乱数生成と呼ぶ．本稿では，ある与えられた定常離散無記憶情報源から出力される固定長のコイン乱数列を変換して，決められた定常離散無記憶情報源から出力されるターゲット乱数列を近似するという乱数生成問題を考える．変換の方法として，情報源符号化の手法に基づく2つのアルゴリズムを提案する．各々のアルゴリズムについて，出力系列長に対する入力系列長の比率がコイン乱数列を出力する情報源のエントロピーに対するターゲット乱数列を出力する情報源のエントロピーの比率を越えない場合，変動距離で測る乱数列の近似誤差が系列長の指数関数のオーダーで0に収束することを示し，その指数を陽に計算する．また，上述の系列長の比率が，エントロピーの比率を越える場合，どんな変換を用いても，変動距離は，系列長を大きくするとその指数関数のオーダーで2に収束することを示し，その指数を陽に計算する．

## 1 まえがき

ある与えられた情報源からの出力系列を変換して，所望の情報源からの系列の出力プロセスを実現あるいは近似することを乱数生成と呼ぶ．乱数生成問題は，計算機科学の分野でにおいて古くから研究されてきた．特に Elias [1], Knuth and Yao [2] の論文において情報理論と乱数生成問題との間の興味深い関係が見られる．1993年の Han and Verdú [3] によって乱数生成問題の本質が明らかにされ，情報理論の立場から，その一般理論が構築された．その後，乱数生成問題の一般理論に関する研究が，Vembu and Verdú [4], Steinberg and Verdú [5], [6], 韓 [7] らにより行なわれ，理論の整備，精密化が進んでいる．

Han and Verdú [3] に始まる上記の一連の研究は，いずれも乱数生成問題の一般理論に関するものであり，計算効率や装置化に注目した実用的な乱数生成アルゴリズムについての議論は十分ではなかった．最近，具体的な乱数生成アルゴリズムの提案とその性能解析に関する研究が，Han and Hoshi [8], 金谷 [9], Uyematsu and Kanaya [10], 大濱 [11], [12] ( Oohama [13]) らにより行なわれ，この方面の研究においても幾つかの進展をみるようになってきた．

乱数生成において，与えられた情報源からの出力系列をコイン乱数列，所望の情報源からの出力系列をターゲット乱数列と呼ぶ．Han and Hoshi [8] は，与えられた定常離散無記憶情報源から出力される可変長 (Variable Length) のコイン乱数列を変換して，ある規定された定常離散無記憶情報源から出力される固定長 (Fixed Length) の ターゲット乱数列を作り出すという，いわゆる Variable to Fixed (VF) 型の乱数生成問題を考察した．彼らは，区間アルゴリズムと呼ばれる単純な乱数生成アルゴリズムを提案し，ターゲット

乱数列を作りだすのに必要な入，出力系列の平均符号長の比が，コイン，ターゲット乱数列を出力する情報源の2つのエントロピーの比で特徴付けられることを証明した．

一方，Uyematsu and Kanaya [10] は，与えられた定常離散無記憶情報源から出力される固定長のコイン乱数列を変換して，指定された定常離散無記憶情報源から出力される固定長のターゲット乱数列を近似するというFF型の乱数生成問題を扱い，特にコイン乱数列が一様分布に従って発生する場合に注目した．この場合の乱数生成問題は，ターゲット乱数列を一般化情報源からの出力系列であるとするかなり一般的な枠組の下で Han and Verdú [3] により考察されており，一般に Resolvability 問題と呼ばれている．Uyematsu and Kanaya [10] は Han and Hoshi [8] の区間アルゴリズムを少し変更して固定長の場合にも適用できるようにしたものを提案している．さらに提案アルゴリズムにより得られる近似乱数とターゲット乱数の変動距離の漸近的ふるまいを解析し，幾つかの結果を得ている．

また，大濱 [11]（ Oohama [13]）は，上述のFF型乱数生成問題について，特にターゲット乱数列が一様分布に従って発生する場合を研究した．これは，Uyematsu and Kanaya [10] の扱った問題のいわば双対問題である．この問題は，コイン乱数列を一般化情報源からの出力系列であるとする一般的な枠組の下で Vembu and Verdú [3] により研究され，Intrinsic Randomness 問題と呼ばれている．大濱は，算術符号における符号化，復号化のプロセスを直接利用した簡単な乱数生成アルゴリズム（以後これを算術アルゴリズムと呼ぶことにする）を提案し，これが Resolvability 問題において区間アルゴリズムと同等の性能を有することを証明した．さらに Intrinsic Randomness 問題に，提案アルゴリズムを適用した場合の性能解析を行ない，幾つかの結果を導いている．その後，大濱 [12]（ Oohama [13]）は，別な乱数生成アルゴリズムとして，文字列の生起確率に基づくソートを用いたFF型乱数生成アルゴリズムを提案している．これを以後ソートアルゴリズムと呼ぶことにする．ソートアルゴリズムは，情報源符号化における Shannon and Fano 符号化法に似た形をしている．大濱 [12]（Oohama [13]）はソートアルゴリズムの性能解析を行ない，Resolvability 問題において，これが算術アルゴリズムと同等の性能を有することを証明した．また，Intrinsic Randomness 問題において，ソートアルゴリズムは，算術アルゴリズムよりもよい性能をもつことを示した．

本稿では，Uyematsu and Kanaya [10]，大濱 [11] の扱ったFF型乱数生成問題のより一般の場合として，コイン乱数列，ターゲット乱数列の従う分布に共に偏りがある場合を考察する．まず，算術アルゴリズムを適用した場合の性能解析を行ない，出力系列長に対する入力系列長の比率がコイン乱数列を出力する情報源のエントロピーに対するターゲット乱数列を出力する情報源のエントロピーの比率を越えない場合，変動距離で測る乱数列の近似誤差が系列長の指数関数のオーダーで0に収束することを示し，その指数を陽に計算する．次にソートアルゴリズムを適用した場合の性能解析を行ない，変動距離の0に近づく指数において，ソートアルゴリズムは算術アルゴリズムよりも良い結果を与えることを証明する．また，上述した系列長の比率がエントロピーの比率を越える場合，どんな変換器を用いても，変動距離は，系列長の指数関数のオーダーで2に収束することを示し，その指数を陽に計算する．

本稿で得られた結果は，Resolvability 問題および Intrinsic Randomness 問題につい

て，各々 Uyematsu and Kanaya [10], 大濱 [11], [12]（Oohama [13]）および大濱 [11], [12]（Oohama [13]）が得た結果を特別な場合として含んでいる．

## 2 乱数生成問題の基本的枠組

乱数生成とは，ある与えられた情報源からの出力系列を変換して，所望の情報源からの系列の出力プロセスを実現あるいは近似することである．与えられた情報源からの出力系列をコイン乱数列，所望の情報源からの出力系列をターゲット乱数列と呼ぶ．また変換器を乱数生成器と呼ぶ．本稿では，コイン乱数列およびターゲット乱数列が，いずれも定常離散無記憶情報源からの出力列である場合を考える．

$X, Y$を各々有限集合 $\mathcal{X}=\{0,1,\cdots,|\mathcal{X}|-1\}$, $\mathcal{Y}=\{0,1,\cdots,|\mathcal{Y}|-1\}$ に値をとる確率変数とし，その確率分布を各々$P_X=\{P_X(x)\}_{x\in\mathcal{X}}$, $P_Y=\{P_Y(y)\}_{y\in\mathcal{Y}}$とする．$\mathcal{X},\mathcal{Y}$ 上の確率分布全体からなる集合を各々 $\mathcal{P}(\mathcal{X}),\mathcal{P}(\mathcal{Y})$ とする．$P=\{P(x)\}_{x\in\mathcal{X}}\in\mathcal{P}(\mathcal{X})$ に対し，$H(P)$ は，確率分布 $P$ から計算されるエントロピー

$$H(P)=\sum_{x\in\mathcal{X}}-P(x)\log P(x) \tag{1}$$

を表す．また $P,P'\in\mathcal{P}(\mathcal{X})$ に対し，$D(P||P')$ は $P$ と $P'$の間のダイバージェンス

$$D(P||P')=\sum_{x\in\mathcal{X}}P(x)\log\left(\frac{P(x)}{P'(x)}\right) \tag{2}$$

を表すものとする．本稿を通じて対数の底は 2 とする．

確率変数$X,Y$によって指定される定常離散無記憶情報源を各々 $\{X_t\}_{t=1}^{\infty}$, $\{Y_t\}_{t=1}^{\infty}$ とし，情報源から発生する各々長さ $n$, $m$ の確率変数の列を $X^n=X_1X_2\cdots X_n$, $Y^m=Y_1Y_2\cdots Y_m$とする．また，出力文字列を $x^n=x_1x_2\cdots x_n\in\mathcal{X}^n$, $y^m=y_1y_2\cdots y_m\in\mathcal{Y}^m$ と表す．$X^n$, $Y^m$の分布を各々$P_X^n=\{P_X^n(x^n)\}_{x^n\in\mathcal{X}^n}$, $P_Y^m=\{P_Y^m(y^m)\}_{y^m\in\mathcal{Y}^m}$と記す．

次に本稿で議論する乱数生成問題の基本的枠組について説明する．ターゲット乱数列，コイン乱数列が各々確率変数 $X$, $Y$によって指定される定常離散無記憶情報源からの出力列である場合を考える．$X$をターゲット乱数，$Y$をコイン乱数と呼ぶ．乱数生成器は，写像 $\varphi^{(n)}:\mathcal{Y}^m\to\mathcal{X}^n$ によって定義され，写像 $\varphi^{(n)}$ は変換レート制約 $m\le nr$ を満たすとする．与えられた$r>0$ に対し，このような変換レート制約を満たす写像 $\varphi^{(n)}$全体からなる集合を$\Phi_n(r)$ とする．$X^n$と$\hat{X}^n$ との一致の尺度として，次に示す変動距離を用いる．

$$d(\hat{X}^n,X^n)=\sum_{x^n\in\mathcal{X}^n}|P_{\hat{X}^n}(x^n)-P_X^n(x^n)|.$$

韓 [7] による乱数問題に対する一般論より直ちに次の 2 つの結果が得られる．

**定理 1** (韓 [7]) $r>H(P_X)/H(P_Y)$ ならば，写像の列 $\{\varphi^{(n)}:\varphi^{(n)}\in\Phi_n(r)\}_{n=1}^{\infty}$ が存在して，

$$\lim_{n\to\infty}d(\varphi^{(n)}(Y^m),X^n)=0.$$

**定理 2** (韓 [7]) $r < H(P_X)/H(P_Y)$ ならば，任意の写像の列 $\{\varphi^{(n)} : \varphi^{(n)} \in \Phi_n(r)\}_{n=1}^{\infty}$ に対し，

$$\lim_{n\to\infty} d(\varphi^{(n)}(Y^m), X^n) = 2.$$

そこで，変動距離が 0 あるいは 2 に近づく速さを調べるために

$$\alpha_n(r, P_X, P_Y) = \min_{\varphi^{(n)}\in\Phi_n(r)} d(\varphi^{(n)}(Y^m), X^n)$$

$$E(r, P_X, P_Y) = \lim_{n\to\infty}\left(-\frac{1}{n}\right)\log\alpha_n(r, P_X, P_Y)$$

$$C(r, P_X, P_Y) = \lim_{n\to\infty}\left(-\frac{1}{n}\right)\log\{2 - \alpha_n(r, P_X, P_Y)\}$$

とおく．上記の変動距離指数の陽な計算公式を与えることが問題である．

## 2.1 幾つかの関数の定義とその性質

ここでは，変動距離指数に関する結果を記述するために必要な幾つかの関数の定義とその性質について記す．

**定義 1**

$$F_\lambda(R, P_X) = \min_{P\in\mathcal{P}(\mathcal{X})}\left\{[\lambda(R - H(P) - D(P||P_X))]^+ + D(P||P_X)\right\}$$

と定義する．ここで $\lambda \in (-\infty, \infty)$，$[t]^+ = \max\{0, t\}$ である．

**定義 2** 確率分布 $P_{X,\lambda} = \{P_{X,\lambda}(x)\}_{x\in\mathcal{X}}$ を $P \in \mathcal{P}(\mathcal{X})$ の関数

$$\lambda(R - H(P) - D(P||P_X)) + D(P||P_X)$$

の最小値を与える分布とする．この様な $P_{X,\lambda}$ は一意的に決まり，それは

$$P_{X,\lambda} = \left\{\frac{P_X^{1-\lambda}(x)}{\sum_{x\in\mathcal{X}} P_X^{1-\lambda}(x)}\right\}_{x\in\mathcal{X}} \tag{3}$$

で与えられることを容易に証明できる．次に

$$R_\lambda(P_X) = H(P_{X,\lambda}) + D(P_{X,\lambda}||P_X),$$
$$R_+(P_X) = \lim_{\lambda\to+\infty} R_\lambda(P_X),$$
$$R_-(P_X) = \lim_{\lambda\to-\infty} R_\lambda(P_X)$$

とおき，さらに

$$F_+(R, P_X) = \lim_{\lambda\to+\infty} F_\lambda(R, P_X),$$
$$F_-(R, P_X) = \lim_{\lambda\to-\infty} F_\lambda(R, P_X)$$

とおく．以上のように定義された関数は次の性質を持つ．

**性質 1** a)

$$\left.\begin{aligned} R_+(P_X) &= \max_{x\in\mathcal{X}}(-\log P_X(x)), \\ R_-(P_X) &= \min_{x\in\mathcal{X}}(-\log P_X(x)) . \end{aligned}\right\} \tag{4}$$

b) $0 \le R \le R_+(P_X)$ に対し，関数 $F_+(R, P_X)$ は $R$の関数として非負，単調増加かつ下に凸であり，

$$F_+(R, P_X) = \min_{\substack{P\in\mathcal{P}(\mathcal{X}):\\ H(P)+D(P\|P_X)\ge R}} D(P\|P_X) \tag{5}$$

の形をもつ．また $0 \le R \le H(P_X)$ のときそのときに限り零になる．

c) $R \ge R_-(P_X)$ に対し，$F_-(R, P_X)$ は $R$の関数として非負，単調減少かつ下に凸であり

$$F_-(R, P_X) = \min_{\substack{P\in\mathcal{P}(\mathcal{X}):\\ H(P)+D(P\|P_X)\le R}} D(P\|P_X) \tag{6}$$

の形をもつ．また $R \ge H(P_X)$ のときそのときに限り零になる．

**性質 2** a) $\lambda \in [0, +\infty)$ に対し

$$F_\lambda(R, P_X) = \begin{cases} F_+(R, P_X) \\ \quad \text{for } H(P_X) < R \le R_\lambda(P_X), \\ \lambda(R - R_\lambda(P_X)) + F_+(R_\lambda(P_X), P_X) \\ \quad \text{for } R > R_\lambda(P_X) \end{cases} \tag{7}$$

が成り立つ．

b) $\lambda \in (-\infty, 0]$ に対し

$$F_\lambda(R, P_X) = \begin{cases} F_-(R, P_X) \\ \quad \text{for } R_\lambda(P_X) \le R \le H(P_X), \\ \lambda(R - R_\lambda(P_X)) + F_-(R_\lambda(P_X), P_X) \\ \quad \text{for } R < R_\lambda(P_X) \end{cases} \tag{8}$$

が成り立つ．

c) 関数 $F_\lambda(R, P_X)$ は，以下のような最適化問題として表現できる．即ち $\lambda > 0$ のとき

$$\begin{aligned} &F_\lambda(R, P_X) \\ &= \min_{0\le\tilde{R}\le R_+(P_X)} \left\{ [\lambda(R - \tilde{R})]^+ + F_+(\tilde{R}, P_X) \right\} \end{aligned} \tag{9}$$

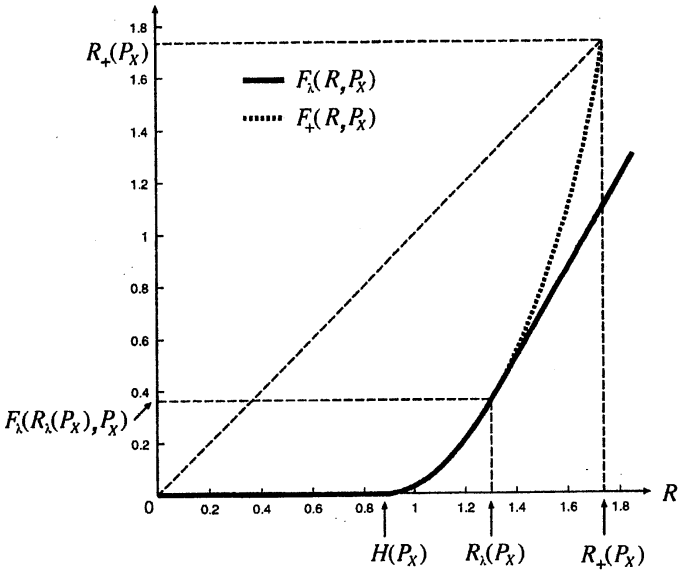


図 1 関数 $F_+(R,P_X)$ および $F_\lambda(R,P_X)$, $\lambda = 1.7$ の形

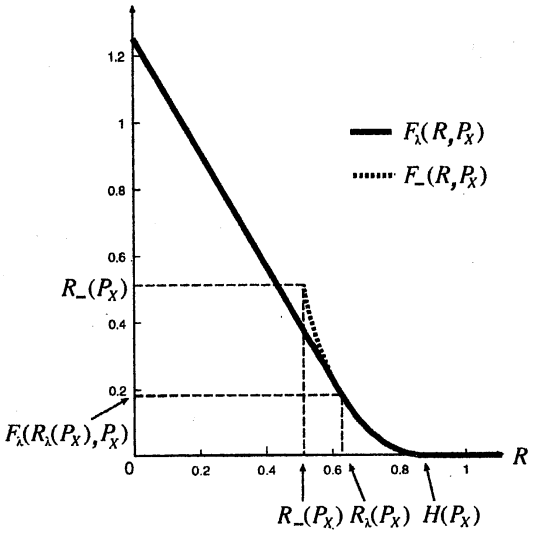


図 2 関数 $F_-(R,P_X)$ および $F_\lambda(R,P_X)$, $\lambda = -1.7$ の形

が成り立つ．また $\lambda < 0$ のとき

$$F_\lambda(R,P_X) = \min_{\tilde{R}\geq R_-(P_X)} \left\{ [\lambda(R-\tilde{R})]^+ + F_-(\tilde{R},P_X) \right\} \tag{10}$$

が成り立つ．

確率変数 $X$ の分布 $P_X$が

$$\mathcal{X} = \{0,1\}\ , P_X(0) = 0.3 \tag{11}$$

で与えられる場合を考える．この場合の関数 $F_+(R,P_X)$ および $F_1(R,P_X), \lambda = 1.7$ の形を図 1 に示す．また，同じ場合における関数 $F_-(R,P_X)$ および $F_\lambda(R,P_X), \lambda = -1.7$ の形を図 2 に示す．

**定義 3**

$$G_+(R,P_X) = \min_{\substack{P\in\mathcal{P}(\mathcal{X}):\\ H(P)\geq R}} D(P||P_X)\,,$$
$$G_-(R,P_X) = \min_{\substack{P\in\mathcal{P}(\mathcal{X}):\\ H(P)\leq R}} D(P||P_X)$$

と定義する．

**性質 3** a) 関数 $G_+(R,P_X)$ は $0 \leq R \leq \log|\mathcal{X}|$ に対し，$R$の関数として，非負で単調増加かつ下に凸 であり，また $0 \leq R \leq H(P_X)$ のときそのときに限り零になる．

b) 関数 $G_-(R,P_X)$ は $R \geq 0$ に対し，$R$の関数として，非負で単調減少かつ下に凸であり，また $R \geq H(P_X)$ のときそのときに限り零になる．

関数 $G_+(R,P_X)$ と $F_+(R,P_X)$ および関数 $G_-(R,P_X)$ と $F_-(R,P_X)$ との間には，各々次のような関係が成り立つ．

**性質 4** a) $0 \leq R \leq H(P_X)$ のとき

$$G_-(R,P_X) = F_-(R+G_-(R,P_X),P_X) \tag{12}$$

が成り立つ．

b) $H(P_X) \leq R \leq \log|\mathcal{X}|$ のとき

$$G_+(R,P_X) = F_+(R+G_+(R,P_X),P_X) \tag{13}$$

が成り立つ．また $\log|\mathcal{X}| \leq R \leq R_+(P_X) - G_+(\log|\mathcal{X}|,P_X)$ のとき

$$G_+(\log|\mathcal{X}|,P_X) \leq F_1(R+G_+(\log|\mathcal{X}|,P_X),P_X) \tag{14}$$

が成り立つ．

# 3 乱数生成アルゴリズムとその性能解析

本章では，2つの乱数生成アルゴリズムを提案し，その性能解析に関して得られる結果を述べる．また，変動距離が2に近付く速さを示す指数 $C(r,P_X,P_Y)$ の下界に関して得られる結果を述べる．

## 3.1 算術アルゴリズム

本節では，情報源符号化における算術符号法のアルゴリズムをそのまま利用する算術アルゴリズムと呼ばれる簡単な乱数生成アルゴリズムを提案し，この性能解析において得られる結果を述べる．区間 $I=[0,1)$ を考える．分布 $P_X$ に基づく累積確率を

$$S_X(0) = 0, \tag{15}$$

$$S_X(x) = \sum_{i<x} P_X(i), 1 \leq x \leq |\mathcal{X}|-1 \tag{16}$$

と定義し，これらを用いて区間 $I$ の分割を

$$I_X(x) = [S_X(x), S_X(x)+P_X(x)) \tag{17}$$

と定義する．さらに写像 $\tau_X : I \to I$ および量子化写像 $\phi_X : I \to \mathcal{X}$ を

$$\tau_X(z) = (P_X(x))^{-1}(z - S_X(x)), \text{ if } z \in I_X(x), \tag{18}$$

$$\phi_X(z) = x, \text{ if } z \in I_X(x) \tag{19}$$

と定義する．適当な初期値 $z$ から出発して，写像の合成と量子化写像により得られる記号列 $\phi_X(z)\phi_X(\tau_X(z))\cdots\phi_X(\tau_X^{k-1}(z))$ を乱数系列として用いることを考えよう．任意に与えられた $k$ と任意に与えられた長さ $k$ の記号列 $x^k = x_1x_2\cdots x_k \in \mathcal{X}^k$ に対し，これを出力するような初期値 $z$ の集合を $I_X(x^k)$ と定義する．これは，$I_X(x^k) = [L_X(x^k), U_X(x^k))$ なる形の半開区間になり，その下端 $L_X(x^k)$ および上端 $U_X(x^k)$ は

$$L_X(x_1) = S_X(x_1),\ U_X(x_1) = S_X(x_1) + P_X(x_1) \tag{20}$$

$$L_X(x^i) = L_X(x^{i-1}) + P_X^{i-1}(x^{i-1})S_X(x_i), \tag{21}$$

$$U_X(x^i) = L_X(x^i) + P_X^i(x^i),\ \text{for } 2 \le i \le k \tag{22}$$

なる漸化式を満たす．確率変数 $Y$ についても確率変数 $X$ の場合と全く同様の定義，記法を採用する．写像の合成と量子化写像により系列を求める手続きは，算術符号における復号化の手続きそのものであり，また与えられた記号列から区間の上端，下端を求める手続きは算術符号における符号化の手続きに他ならない．

算術アルゴリズムは次のようになる．

**算術アルゴリズム**　変換レートを $r = m/n$ とする．$y^m \in \mathcal{Y}^m$ に対し，写像 $\varphi_{\mathrm{a}}^{(n)} : \mathcal{Y}^m \to \mathcal{X}^n$ を

$$\varphi_{\mathrm{a}}^{(n)}(y^m) = x_1x_2\cdots x_n, \tag{23}$$

$$x_i = \phi_X(\tau_X^{i-1}(L_Y(y^m))),\ \text{for } i = 1, 2, \cdots, n \tag{24}$$

と定義する．

算術アルゴリズムの性能評価に関する結果を述べるために

$$E_{\mathrm{a}}(r, P_X, P_Y) = \min_{0 \le R \le \log|\mathcal{X}|} \max\left\{G_+(R, P_X), rG_-\left(\frac{R}{r}, P_Y\right)\right\}$$

と定義する．

算術アルゴリズムに対する我々の結果は以下のようである．

**定理 3** 任意の $r > 0$ と算術アルゴリズムで定義される写像の列 $\{\varphi_{\mathrm{a}}^{(n)} : \varphi_{\mathrm{a}}^{(n)} \in \Phi_n(r)\}_{n=1}^{\infty}$ に対し

$$\lim_{n\to\infty}\left(-\frac{1}{n}\right)\log d(\varphi_{\mathrm{a}}^{(n)}(Y^m), X^n) \ge E_{\mathrm{a}}(r, P_X, P_Y) \tag{25}$$

が成り立つ．

定理 3 より $r > H(P_X)/H(P_Y)$ のとき，算術アルゴリズムによる近似誤差は出力系列長 $n$ の指数関数のオーダーで 0 に収束し，その指数は $E_{\mathrm{a}}(r, P_X, P_Y)$ で与えられる．

$P_X$ が $\mathcal{X}$ 上の一様分布であるとき，$G_+(R, P_X)$ は，その定義域 $0 \le R \le \log|\mathcal{X}|$ において，定数 0 をとる．この場合 $E_{\mathrm{a}}(R, P_X, P_Y)$ は $rG_-(\frac{\log|\mathcal{X}|}{r}, P_Y)$ と一致する．この下界は，Intrinsic Randomness 問題において，大濱 [11]（Oohama [13]）の得た下界と一致している．したがって，定理 3 は，Intrinsic Randomness 問題において，大濱 [11]（Oohama [13]）が得た結果を特別な場合として含む．

関数 $E_{\mathrm{a}}(r, P_X, P_Y)$ について次の性質が成り立つ．

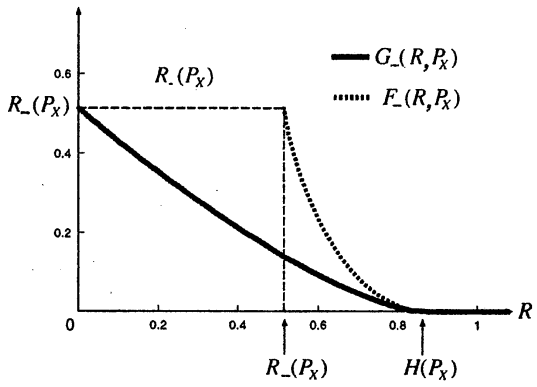


図 3 関数 $G_-(R,P_X)$ および $F_-(R,P_X)$ の形

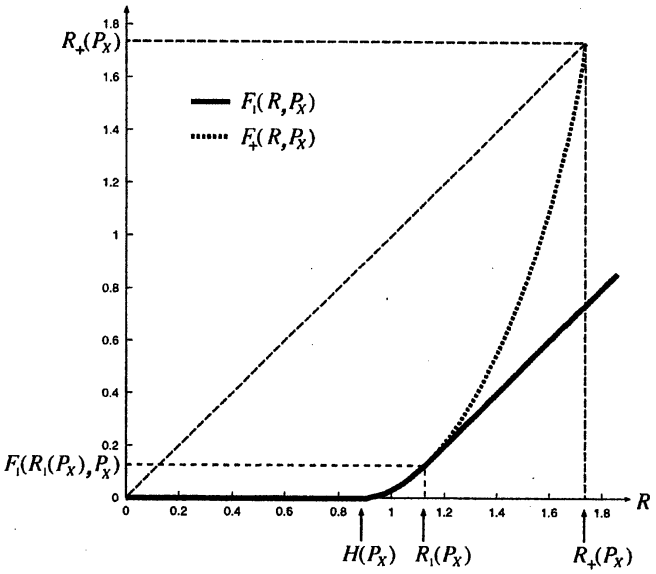


図 4 関数 $F_1(R,P_X)$ および $F_+(R,P_X)$ の形

**性質 5** 関数 $E_{\rm a}(r,P_X,P_Y)$ は以下のような別表現を持つ.

$$E_{\rm a}(r,P_X,P_Y) = \min_{R\geq rR_-(P_Y)} \max\left\{F_1(R,P_X), rF_-\left(\frac{R}{r},P_Y\right)\right\}. \tag{26}$$

$P_Y$ が $\mathcal{Y}$ 上の一様分布であるとき，$F_-(R,P_X)$ は，その定義域 $R\geq r\log|\mathcal{Y}|$ において，定数 0 をとる．この場合 $E_{\rm a}(R,P_X,P_Y)$ は Resolvability 問題における算術アルゴリズムの性能解析において大濱 [11]（Oohama [13]）が得た下界 $F_1(r\log|\mathcal{Y}|,P_X)$ と一致する．したがって，定理 3 は，大濱 [11]（Oohama [13]）が Resolvability 問題において得た結果を特別な場合として含む．(11) で特定される場合における関数 $G_-(R,P_X)$ および $F_-(R,P_X)$ の形を図 3 に示す．また，同じ場合における関数 $F_1(R,P_X)$ および $F_+(R,P_X)$ の形を図 4 に示す．

## 3.2 ソートアルゴリズム

ここでは，ソートアルゴリズムを提案し，それに関して得られる結果について述べる．まず，そのために必要な幾つかの定義と補題について述べる．

**定義 4** 長さ $n$ の任意の系列 $x^n=x_1x_2\cdots x_n\in\mathcal{X}^n$ に対し，$n(x|x^n)$ を $x_i=x$ を満たす $i$ の個数とする．系列の中に現れる文字の相対頻度分布 $\{n(x|x^n)/n\}_{x\in\mathcal{X}}$ を系列 $x^n$ のタイプとよび，$P_{x^n}$ と記す．また $\mathcal{X}$ 上のタイプ全体からなる集合を $\mathcal{P}_n(\mathcal{X})$ と記す．さらに $\hat{P}\in\mathcal{P}_n(\mathcal{X})$ に対し

$$T^n_{\hat{P}} = \left\{x^n \,|\, P_{x^n}=\hat{P}\right\} \tag{27}$$

とおく.

系列のタイプ，タイプの集合について次の補題が成り立つ．証明は Csiszár and Körner [14] を参照せよ.

**補題 1** a) $|\mathcal{P}_n(\mathcal{X})| \leq (n+1)^{|\mathcal{X}|}$.

b) $\hat{P} \in \mathcal{P}_n(\mathcal{X})$ に対し

$$(n+1)^{-|\mathcal{X}|}2^{nH(\hat{P})} \leq |T^n_{\hat{P}}| \leq 2^{nH(\hat{P})}.$$

c) $x^n \in T^n_{\hat{P}}$ に対し

$$P^n_X(x^n) = 2^{-n[H(\hat{P})+D(\hat{P}\|P_X)]}.$$

次にソートアルゴリズムについて述べる．タイプ $\hat{Q} \in \mathcal{P}_m(\mathcal{Y})$ を $T^m_{\hat{Q}}$ の各元の分布 $P^m_Y$ に基づく発生確率 $2^{-m[H(\hat{Q})+D(\hat{Q}\|P_Y)]}$ の順番に並べる．値が同じものがある場合はその中で順番を適当に決める．さらに，タイプの集合 $T^m_{\hat{Q}}$ に属する各元を辞書式の順序に従って並べる．このような順序付けに基づいて $\mathcal{Y}^m$ の系列をその発生確率の大きさの順番に並べることができる．これを

$$P^m_Y(y^m_1) \geq P^m_Y(y^m_2) \geq \cdots \geq P^m_Y(y^m_{|\mathcal{Y}|^m}) \tag{28}$$

とする．また，$y^m \in \mathcal{Y}^m$ に付けられる番号を $i(y^m)$ で表すことにする．区間 $I = [0,1)$ を考える．分布 $P^m_Y$ に基づく累積確率を

$$S_Y(y^m_1) = 0, \tag{29}$$

$$S_Y(y^m) = \sum_{a^m : i(a^m) < i(y^m)} P^m_Y(a^m) \tag{30}$$

と定義し，これらを用いて区間 $I$ の分割を

$$I_Y(y^m) = [S_Y(y^m), S_Y(y^m) + P^m_Y(y^m)) \tag{31}$$

と定義する．確率変数 $X$ についても確率変数 $Y$ の場合と全く同様の定義，記法を採用する.

ソートアルゴリズムは，以下のように述べられる.

**ソートアルゴリズム** 変換レートを $r = m/n$ にとる．長さ $m$ の入力列 $y^m \in \mathcal{Y}^m$ から $L_Y(y^m)$ を計算する．$L_Y(y^m)$ に対し，$I_X(x^n)$ となる $x^n \in \mathcal{X}^n$ が唯一つ存在する．これを用いて写像 $\varphi^{(n)}_s : \mathcal{Y}^m \to \mathcal{X}^n$ を $\varphi^{(n)}_s(y^m) = x^n$ と定める.

ソートアルゴリズムの性能評価に関する結果を述べるために

$$\mathcal{R}_s = \Big\{(R,\tilde{R}) : \ R \geq rR_-(P_Y), \\ rF_-(\tfrac{R}{r}, P_Y) \leq \tilde{R} \leq R_+(P_X)\Big\}$$

とおき，

$$E_s(r, P_X, P_Y) \\ = \min_{(R,\tilde{R})\in\mathcal{R}_s} \Big\{[R-\tilde{R}]^+ \\ + \max\Big\{F_+(\tilde{R}, P_X), rF_-\Big(\frac{R}{r}, P_Y\Big)\Big\}\Big\}$$

と定義する．ソートアルゴリズムに対する結果は次のようになる.

**定理 4** 任意の $r>0$ とソートアルゴリズムで定義される写像の列 $\{\varphi_{\mathrm{s}}^{(n)}:\varphi_{\mathrm{s}}^{(n)}\in\Phi_n(r)\}_{n=1}^{\infty}$ に対し

$$\lim_{n\to\infty}\left(-\frac{1}{n}\right)\log d(\varphi_{\mathrm{s}}^{(n)}(Y^m),X^n)\geq E_{\mathrm{s}}(r,P_X,P_Y) \tag{32}$$

が成り立つ.

定理4より $r>H(P_X)/H(P_Y)$ のとき，ソートアルゴリズムによる近似誤差は出力系列長 $n$ の指数関数のオーダーで0に収束し，その指数が $E_{\mathrm{s}}(r,P_X,P_Y)$ で与えられる．この関数は，次の性質に示されるような興味深い下界をもつ.

**性質 6** 任意の $r>0$ に対し

$$E_{\mathrm{s}}(r,P_X,P_Y)\geq E_{\mathrm{a}}(r,P_X,P_Y) \tag{33}$$

が成り立つ.

上記の性質はソートアルゴリズムが算術アルゴリズムよりも良い性能を持つことを意味している．$P_Y$ が $\mathcal{Y}$ 上の一様分布であるとき，$E_{\mathrm{s}}(R,P_X,P_Y)$ は Resolvability 問題におけるソートアルゴリズムの性能解析において大濱 [12]（Oohama [13]）が得た下界 $F_1(r\log|\mathcal{Y}|,P_X)$ と一致する．したがって定理4は，Resolvability 問題におけるソートアルゴリズムの性能解析に関して大濱 [12]（Oohama [13]）が得た結果を特別な場合として含む.

つぎに関数 $E_{\mathrm{s}}(r,P_X,P_Y)$ の別な下界を導く．そのために

$$\begin{aligned}&\hat{F}_{-1}(R,P_X)\\&=\min_{\substack{P\in\mathcal{P}(\mathcal{X}):\\ D(P\|P_X)\leq R}}\Big\{[H(P)+D(P\|P_X)-R]^+\\&\qquad+D(P\|P_X)\Big\}\end{aligned}$$

なる関数を定義する．$F_-^{-1}(R,P_X)$, $0<R<R_-(P_X)$ を $R_-(P_X)<R\leq H(P_X)$ における $F_-(R,P_X)$ の逆関数とし，$\hat{R}_{-1}(P_X)=F_-(R_{-1}(P_X),P_X)$ とおく．このとき簡単な計算により以下を示すことができる.

$$\begin{aligned}&\hat{F}_{-1}(R,P_X)\\&=\begin{cases}F_-^{-1}(R,P_X)\\\quad\text{for } 0<R\leq\hat{R}_{-1}(P_X),\\-R+R_{-1}(P_X)+\hat{R}_{-1}(P_X)\\\quad\text{for } \hat{R}_{-1}(P_X)\leq R\leq R_{-1}(P_X),\\F_-(R,P_X)\\\quad\text{for } R\geq R_{-1}(P_X).\end{cases}\end{aligned} \tag{34}$$

(11) で指定される場合における関数 $\hat{F}_{-1}(R,P_X)$ の形を図5に示す．上記のように定義さ

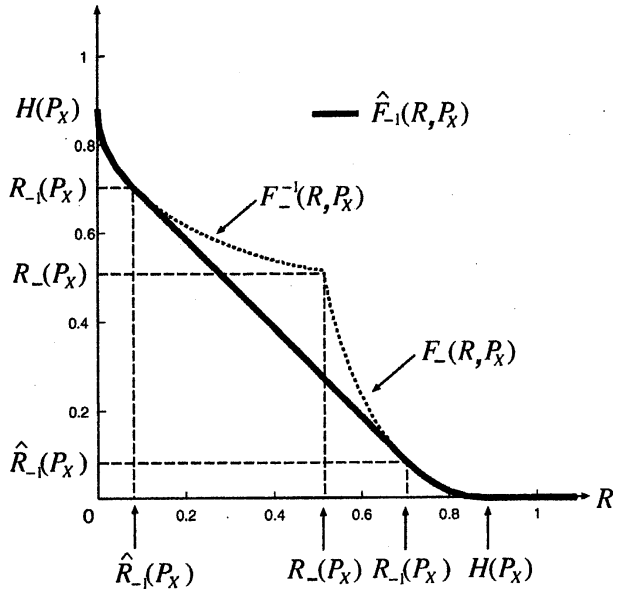


図 5 関数 $\hat{F}_{-1}(R, P_X)$ の形

れた関数 $\hat{F}_{-1}(\cdot, P_X)$ を用いて，関数 $\hat{E}_s(\ r, P_X, P_Y)$ を

$$\begin{aligned}&\hat{E}_s(r, P_X, P_Y)\\&= \min_{0\le\tilde{R}\le R_+(P_X)} \max\left\{F_+(\tilde{R}, P_X), r\hat{F}_{-1}\left(\frac{\tilde{R}}{r}, P_Y\right)\right\}\end{aligned}$$

と定義する．このとき $\hat{F}_{-1}(\tilde{R}, P_Y)$ が以下の最適化問題の形の別表現

$$\begin{aligned}&\hat{F}_{-1}(\tilde{R}, P_Y)\\&= \min_{\substack{R\ge R_-(P_Y),\\ F_-(R,P_Y)\le\tilde{R}}} \left\{[R-\tilde{R}]^+ + F_-(R, P_Y)\right\} \qquad (35)\\&= \min_{R\ge\hat{F}_-^{-1}(\tilde{R},P_Y)} \left\{[R-\tilde{R}]^+ + F_-(R, P_Y)\right\} \qquad (36)\end{aligned}$$

をもつことに注目すれば，これと関数 $E_s(r, P_X, P_Y)$ の定義から，以下の性質が成り立つことを証明できる．

**性質 7** 任意の $r > 0$ に対し

$$E_s(r, P_X, P_Y) \ge \hat{E}_s(r, P_X, P_Y)\,. \qquad (37)$$

確率分布 $P_X$ が $\mathcal{X}$ 上の一様分布であるとき，関数 $F_+(R, P_X)$ は，定義域 $0 \le R \le \log|\mathcal{X}|$ 上で定数 0 をとる．このとき，関数 $E_s(r, P_X, P_Y)$ の定義から $P_X$ が $\mathcal{X}$ 上の一様分布であるとき，関数 $E_s(r, P_X, P_Y)$ および $\hat{E}_s(r, P_X, P_Y)$ は，Oohama [13] が，Intrinsic Randomeness

問題におけるソートアルゴリズムの解析において得た下界 $r\hat{F}_{-1}\left(\frac{\log|\mathcal{X}|}{r},\ P_Y\right)$ に一致する．よって，定理 4 は，Oohama [13] が Intrinsic Randomeness 問題におけるソートアルゴリズムの性能解析に関して得た結果を特別な場合として含む．

関数 $E_{\mathrm{a}}(r,P_X,P_Y)$ および $E_{\mathrm{s}}(r,P_X,P_Y)$ はともに最適な近似誤差指数関数 $E(r,P_X,P_Y)$ の下界を与える．Resolvability 問題および Intrinsic Randomness 問題に対する結果によれば $P_X$ あるいは $P_Y$ が一様分布の場合は，$E(r,P_X,P_Y)$ の上界が陽に計算されており，それは $r$のある範囲の値において，タイトになる．しかしながら 一般の $P_X$ および $P_Y$ に対する $E(r,P_X,P_Y)$ の陽な上界は今のところ得られていない．

### 3.3 乱数生成問題の強逆定理

乱数生成問題の強逆定理に関する結果を述べるために

$$C_{\mathrm{I}}(r,P_X,P_Y)$$
$$=\min_{0\le R\le\log|\mathcal{X}|}\max\left\{G_-(R,P_X),rG_+\left(\frac{R}{r},P_Y\right)\right\}$$

とおく．このとき次が成り立つ．

**定理 5** 任意の $r>0$ と任意の写像の列 $\{\varphi^{(n)}:\varphi^{(n)}\in\Phi_n(r)\}_{n=1}^{\infty}$ に対し

$$\lim_{n-\infty}\left(-\frac{1}{n}\right)\log\left\{2-d(\varphi^{(n)}(Y^m),X^n)\right\}$$
$$\ge C_{\mathrm{I}}(r,P_X,P_Y) \tag{38}$$

が成り立つ．

定理 5 より $r<H(P_X)/H(P_Y)$ のとき，任意の乱数生成器に対し近似誤差は，出力系列長の指数関数のオーダーで 2 に収束し，その指数の値は関数 $C_{\mathrm{I}}(r,P_X,P_Y)$ の値を下回らないことが分かる．確率分布 $P_Y$ を $\mathcal{Y}$上の一様分布とするとき，関数 $C_{\mathrm{I}}(r,P_X,P_Y)$ は Uyematsu and Kanaya [10] が Resolvability 問題において得た最適な指数と一致する．同様に $P_X$ を $\mathcal{X}$上の一様分布とすれば，$C_{\mathrm{I}}(r,P_X,P_Y)$ は大濱 [11]（Oohama [13]） が Intrinsic Randomness 問題において得た最適な指数と一致する．これらのことから定理 5 は，Resolvability および Intrinsic Randomness 問題に対するこれまでの結果を特別な場合として含むことがわかる．一般の確率分布 $P_X$ および $P_Y$ に対する関数 $C(r,P_X,P_Y)$ の陽な上界 は未だに求められていない．

## 4 むすび

定常離散無記憶情報源から出力される固定長のコイン乱数列を変換して，決められた定常離散無記憶情報源から出力されるターゲット乱数列を近似するという乱数生成問題を考察し，2 つの変換アルゴリズムを提案した．また，提案する 2 つの変換アルゴリズムについて，変動距離で測る近似誤差を評価し，変換効率がエントロピーの比よりも小さい場

合において，誤差が0に近付く速さを示す指数の下界を導いた．また，変換率が大きいところでは，どんな変換を行なっても近似誤差が2に近付くことを示し，その速さを示す指数の下界を導いた．

変動距離指数の上界についての議論，より一般の情報源の場合への拡張が今後の課題として挙げられる．

# 参考文献


[1] P. Elias, "The efficient construction of an unbiased random sequences," *Ann. Math. Statist.*, vol. 43, pp. 865-870, 1972.

[2] D. Knuth and A. Yao, "The complexity of nonuniform random number generation," *Algorithm and Complexity, New Directions and Results,* pp. 357-428, ed. by J. F. Traub, Academic Press, New York, 1976.

[3] T. S. Han and S. Verdú, "Approximation theory of output statistics," *IEEE Trans. Inform. Theory*, vol. 39, pp. 752-722, May 1993.

[4] S. Vembu and S. Verdú, "Generating random bits from an arbitrary source: Fundamental limits," *IEEE Trans. Inform. Theory*, vol. 41, pp. 1322-1332, Sept. 1995.

[5] Y. Steinberg and S. Verdú, "Simulation of random processes and rate-distotion theory," *IEEE Trans. Inform. Theory*, vol. 42, pp. 63-86, Jan. 1996.

[6] Y. Steinberg and S. Verdú, "Channel simulation and coding with side information," *IEEE Trans. Inform. Theory*, vol. 40, pp. 634-646, May 1996.

[7] 韓太舜, 情報理論における情報スペクトル的方法, 培風館, 1998.

[8] T. S. Han and M. Hoshi, "Interval algorithm for random number generation," *IEEE Trans. Inform. Theory*, vol. 43, pp. 599-611, March 1997.

[9] 金谷文夫, "漸近最良なマルコフ乱数発生アルゴリズム," 第20回情報理論とその応用シンポジウム予稿集, pp.77-80, Dec. 1997.

[10] T. Uyematsu and F. Kanaya, "Channel simulation by interval algorithm: A performance analysis for interval algorithm," *IEEE Trans. Inform. Theory*, vol. 45, pp. 2121-2129, Sept. 1999.

[11] 大濱靖匡, "一次元区分線形写像を用いたFF型乱数生成アルゴリズム," 第21回情報理論とその応用シンポジウム予稿集, pp. 57-60, Dec. 1998.

[12] 大濱靖匡, "定常離散無記憶情報源に対する乱数生成問題について," 電子情報通信学会技術研究報告, IT-98-59, pp.25-30, Jan. 1999.

[13] Y. Oohama, "Arithmetic and sorting algorithm for fixed to fixed random number generation and their performance analysis," *preprint.*

[14] I. Csiszár and J. Körner, *Information Theory : Coding Theorems for Discrete Memoryless Systems,* Academic Press, New York, 1981.