Vol.2014-HPC-144 No.1
2014/5/26

生成可能な乱数の精度を保証する離散的な逆変換法

石田 翔太郎[1]　須田 礼仁[1]

概要：高次元数値積分におけるモンテカルロ法や、各種シミュレーションにおける非決定的要素の実現など、計算機上においては実に多くの場面で乱数が使用されてきた。また、特にシミュレーションにおいては、必要とされる乱数が単純な一様乱数ばかりではなく、様々な (時として複雑な) 確率密度関数に従う乱数であることも多い。このような背景から、これまで多くの論文において、生成の容易な一様乱数から様々な確率密度関数に従う乱数への変換法が論じられてきた。
しかしながら、計算機上で乱数を生成する場合、本来連続的である分布を離散的に近似する必要がある。そのため、乱数の変換方法と同時に、離散化の方法や精度を考える必要がある。さもなくば、確率密度関数が谷になっている箇所や、テール領域など、本来の分布が持っていた特色が失われてしまう。特にテール領域に関しては、いくら確率が低くとも、絶対値の大きい値で構成されていることから、計算結果に大きな影響力を持ちうる箇所となる。
これまでにもテールに関して、正規分布のテール領域の精度を上げる取り組みや、テール領域を他の分布で近似すること無く高速にテール領域の乱数を生成する取り組みがあった。しかしながら、それらは精度を保証するものではなかった。そこで、この論文では、テールの精度を向上させるだけでなく、その保証をすることを目指す。その方法として、離散化後の確率分布と離散化の間隔に着目した、反復逆変換法という新しい乱数変換法を挙げ、その正当性を示す。

キーワード：乱数, 離散化, テール, 精度保証, 逆変換法

Shotaro Ishida[1]　Reiji Suda[1]

1. 序論

1.1 背景

高次元数値積分におけるモンテカルロ法や、各種シミュレーションにおける非決定的要素の実現など、計算機上においては実に多くの場面で乱数が使用されてきた。また、特にシミュレーションにおいては、必要とされる乱数が単純な一様乱数ばかりではなく、様々な (時として複雑な) 確率密度関数に従う乱数であることも多い。このような背景から、これまで多くの論文において、生成の容易な一様乱数 [1–3] から様々な確率密度関数に従う乱数への変換法が論じられてきた。正規分布に従う乱数 (正規乱数) を例にいくつか見てみると、二つの互いに独立な一様乱数の組から、同じく二つの互いに独立な正規乱数の組を生成する Box-Muller 法 [4]。Box-Muller 法で必要だった三角関数の計算が不要になる Polar 法 [5,6]。三角分布による区分線形近似を用いた Kabal による方法 [7]。互いに同面積な長方形を用いて、正規分布の確率密度関数を横軸に沿って分割近似する Ziggurat 法 [8]。一方、確率密度関数をパズルのように分割して、一つの長方形に埋め込む Monty-Python 法 [9]。それ以外にも、棄却法によるもの [10] や奇偶法によるもの [11]、二つの一様乱数の比を利用したもの [12] まで存在する。このように、正規乱数を生成する方法だけでも、多種多様な方法が論じられてきた。

しかしながら、計算機上で乱数を生成する場合、本来連続的である分布を離散的に近似する必要がある。そのため、乱数の変換方法と同時に、離散化の方法 [13] や精度を考える必要がある。さもなくば、確率密度関数が谷になっている箇所や、テール領域など、本来の分布が持っていた特色

[1] 東京大学大学院情報理工学系研究科

が失われてしまう。特にテール領域に関しては、いくら確率が低くとも、絶対値の大きい値で構成されていることから、計算結果に大きな影響力を持ちうる箇所となる。これまでにもテールに関して、正規分布のテール領域の精度を上げる取り組み [14] や、テール領域を他の分布で近似すること無く高速にテール領域の乱数を生成する取り組み [15] があった。しかしながら、それらは精度を保証するものではなかった。そこで、この論文では、テールの精度を向上させるだけでなく、その保証をすることを目指す。

1.2 記法

ここでは、記法に関する説明を行う。

- RBG(Random Bit Generator)
  ランダムかつ等確率に 1 ビット生成するもの。
- RDG(Random Discrete-Number Generator)
  離散乱数生成器
- $f : \mathbb{R} \to \mathbb{R}$
  確率密度関数
  F が与えられていれば、f は
  $$f(r) = \frac{dF(r)}{dr}$$
  という微分で定義される。
  なお、テールを持ついくつかの分布の中には、
  $$\forall r \in \mathbb{R},\, f(r) > 0$$
  を満たすものが存在するため、当論文では、この条件を満たす分布のみを考えていく。もちろん、これはいささか強すぎる制約であるが、これはテールの定義の為というよりはむしろ議論を簡単にするためのものであるので、この論文での議論を少し修正するだけで、議論の一般化が可能であると思われる。
- $F : \mathbb{R} \to \mathbb{R}$
  累積分布関数
  f が与えられていれば、F は
  $$F(r) = \int_{-\infty}^{r} f(r)\, dr$$
  という積分で定義される。
  なお、確率密度関数 f の正値性から、F は単調増加関数である。
  また、$0 \leq F(r) \leq 1$ と併せると、F の逆関数 F^{-1} が $[0, 1]$ 上で定義されており、F^{-1} も単調増加関数となる。

1.3 本紙の構成

本紙の構成は、以下の通りである。

- 第一章：序論
  この章である。
- 第二章：取り組み
  テールの精度に関する定義を与えたのち、提案手法のアルゴリズムについて述べる。
- 第三章：正当性
  第二章で述べられたアルゴリズムが、(同じく第二章で定義された) テールの精度を保証していることを実数において証明する。
- 第四章：実験
  アルゴリズムを単純に浮動小数点数で利用した際の挙動について示す。
- 第五章：関連研究
  当研究と関連すると思われる研究に関して紹介する。
- 第六章：結論
  当研究で得られた結論を述べ、今後の課題について提案する。

2. 取り組み

2.1 目的

1.1 章で述べた通り、この論文の目的は、テールの精度を向上させるだけでなく、その保証をすることである。まずこの章においては、テールの精度の指標として、

(1) 乱数生成確率が確率密度関数 f に従うこと

(2) 出現値 (RDG が生成可能な乱数) の間隔が狭いことを順次定義し ((1) は分布の近似精度、(2) は分布の離散化精度のための指標である)、精度向上のための提案する。

なお、以下では、ある RDG によって生成される f に従う乱数の集合を、$\mathbb{DR}_f$(出現値集合と呼ぶ) とおく。

2.2 定義

2.2.1 乱数生成確率

分布を離散化する方法には主に二種類ある [13]。一つは、丸めに基づく方法。もう一つは、モーメント保存に基づく方法である。今回は、簡単のため、丸めに基づく方法を採用した。そのため、乱数生成確率について述べる前に、先に丸めに関して考える。

2.2.1.1 丸め健全

$\mathbb{X}$ を $\mathbb{R}$ の部分集合であるとする。このとき、$r \in \mathbb{R}$ を $x \in \mathbb{X}$ に丸める

$$round_{\mathbb{X}} : \mathbb{R} \to \mathbb{X}$$

が次の条件を満たすとき、この $round_{\mathbb{X}}$ は丸め健全であるという。

丸め健全のための条件

- 全域一意性
  任意の $r \in \mathbb{R}$ に対して、(唯一の) 元 $x \in \mathbb{X}$ が定まり、
  $$round_{\mathbb{X}}(r) = x$$
  となる。
- 同一性
  任意の $x \in \mathbb{X} \subset \mathbb{R}$ に対して、
  $$round_{\mathbb{X}}(x) = x$$
  となる。
- 連続性
  ある $p, q \in \mathbb{R}$ に対して、
  $$round_{\mathbb{X}}(p) = round_{\mathbb{X}}(q) = x$$
  となるならば、任意の $t \in [0,1]$ に対して、
  $$round_{\mathbb{X}}(t * p + (1-t) * q) = x$$
  となる。

2.2.1.2 確率密度関数 f に従う

f を近似するような RDG について考える。この RDG が $d \in \mathbb{DR}_f$ を生成する確率を $P(d)$ とおく。ある丸め健全な $round_{\mathbb{DR}_f}$ が存在して、任意の $d \in \mathbb{DR}_f$ に対して

$$P(d) = \int_{round^{-1}_{\mathbb{DR}_f}[\{d\}]} f(r)\,dr \qquad (1)$$

を満たすとき、この RDG が生成する乱数は、確率密度関数 f に従っていると定義する。ここで、逆像 $round^{-1}_{\mathbb{DR}_f}[\{d\}]$ は以下のように定義される。

$$round^{-1}_{\mathbb{DR}_f}[\{d\}] = \left\{ r \in \mathbb{R} \mid round_{\mathbb{DR}_f}(r) = d \right\}$$

この定義は、もし $\mathbb{R}$ 上で f に従う連続実数乱数を生成できたとして、それを $round_{\mathbb{DR}_f}$ を用いて丸めたものを出力する場合、d を出力する確率 $P(d)$ が、

$$\int_{round^{-1}_{\mathbb{DR}_f}[\{d\}]} f(r)\,dr$$

と等しくなることに由来している。

2.2.2 出現値間隔

出現値集合 $\mathbb{DR}_f$ に対する次の条件を、(b,h) 間隔条件 ($(b,h) \in \mathbb{R} \times \mathbb{R}$) と呼ぶ。(ただし、$0 < h << b$ とする。)

(b,h) 間隔条件

任意の $d \in \mathbb{DR}_f$ に対して、

$$\{d \in \mathbb{DR}_f \mid d - w \le d' < d\}$$
$$\{d \in \mathbb{DR}_f \mid d < d' \le d + w\}$$

がいずれも空集合とならない。ただし、ここでの w は、

$$w = \begin{cases} h & (\text{非正規化領域: } |d| < b) \\ \frac{h}{b}|d| & (\text{正規化領域: } |d| \ge b) \end{cases}$$

とする。

つまり、正規化領域においては相対間隔を一定以下に抑え、非正規化領域においては、絶対間隔を一定以下に抑えるという条件である。

また、d を実数に一般化した次の条件を、(b,h) 離散化条件と呼ぶ。

(b,h) 離散化条件

任意の $r \in \mathbb{R}$ に対して、

$$\{d \in \mathbb{DR}_f \mid r - w \le d < r\}$$
$$\{d \in \mathbb{DR}_f \mid r < d \le r + w\}$$

がいずれも空集合とならない。ただし、ここでの w は、

$$w = \begin{cases} h & (\text{非正規化領域: } |r| < b) \\ \frac{h}{b}|r| & (\text{正規化領域: } |r| \ge b) \end{cases}$$

とする。

なお、自明なことではあるが、(b,h) 離散化条件を満たしていれば、(b,h) 間隔条件も満たしている。

2.3 提案手法

まずは、任意の分布に適用できるように、最も有名で普遍的な手法である、逆変換法を基にした。もちろん、逆変換法には、累積分布関数の逆関数を計算しなくてはならないという制約があるが、逆に言えば、この逆関数さえ計算できるならば、いかなる分布であっても適用可能な方法であると言える。次に、分布の離散化に関して、ユーザーには (b,h) 離散化条件の b と h を要求するのみである。さらには、b と h を指定しなおすことで、離散化の精度をいつでも変更できるようになっている。

2.3.1 アイデア

さて、出現値間隔を抑えるためのアイデアであるが、以下のような状況を考える。

モデル

最初に、ある機械が $[0,1]$ 上の一様乱数 u を実数で生成する。もちろん、本来計算したいのは $F^{-1}(u)$ の値 (r とおく) である。しかしながら、u は無限精度であるため、その値を正確に求めることは出来ない。そのかわりに、u の (二進数表記時における) 小数点以下のビットを順次調べることで、u の近似値をいくらでも正確に求めることができる。

このとき、数ビットを調べた結果、ある $u_{min}, u_{max} \in \mathbb{R}$ に対して、u が

$$u_{min} \leq u \leq u_{max}$$

を満たすことが分かれば、変換結果の乱数 r は、

$$F^{-1}(u_{min}) \leq r \leq F^{-1}(u_{max})$$

を満たすことが分かる。よって、ビットを調べることで u の範囲を制限してゆけば、結果として、r の範囲を狭めることができる。当論文では、このようにして r の範囲を十分狭めた後に $r = F(u)$ の近似値として $F^{-1}(u_{min})$ と $F^{-1}(u_{max})$ の間の値を出力することで、出現値間隔を抑えることを実現している。

2.3.2 RDG 擬似コード

ここで、実際の RDG の擬似コードを示す。まずは準備として、擬似コードの入出力と仕様について、さらに、擬似コードにて使用されている変数について説明する。

2.3.2.1 入出力及び仕様

- 入力
  - 累積分布関数の逆関数 $F^{-1} : \mathbb{R} \to \mathbb{R}$
  - (b,h) 離散化条件を定義するための $(b,h) \in \mathbb{R} \times \mathbb{R}$
- 出力
  - 乱数 $r \in \mathbb{R}$
- 仕様 (計算内容)
  以下の条件を満たす乱数を生成する。
  - 乱数生成確率が (1) 式を満たす。
  - 出現値間隔が (b,h) 離散化条件を満たす。

2.3.2.2 使用変数説明

擬似コードにて使用されている変数の意味は、以下の通りである。

- $n \in \mathbb{N}$
  現在の反復回数。
- $b_k \in \{0,1\}$
  u の範囲を狭めるために、RBG が k 反復目に出力した値。(すなわち、u の小数点以下 k ビット目の値。) ただし、$b_0 = 0$ とする。(これは、u が 1 であるかもしれないことに矛盾するように思えるが、$u = 1$ の時は、小数点以下の全てのビットが 1 であるとき ($u = 0.\dot{1}$) と同一視できる。)
- $u_{min}[k], u_{max}[k] \in \mathbb{R}$
  k 反復目での u_{min} と u_{max} の値。
- $r_{min}, r_{max} \in \mathbb{R}$
  現在の u の範囲に対応する r の範囲、すなわち、

$$r_{min} = F^{-1}(u_{min}[n])$$
$$r_{max} = F^{-1}(u_{max}[n])$$

  で定義される値。

2.3.2.3 擬似コード

実際の擬似コードは、以下の通りである。

00: (初期化)

$$n \leftarrow 0$$
$$b_0 \leftarrow 0$$
$$u_{min}[0] \leftarrow 0.0$$
$$u_{max}[0] \leftarrow 1.0$$

10: (u の範囲を二等分)
$n \leftarrow n+1$(反復回数をインクリメントする。)
RBG によりランダムに 1 ビット生成し、

- 生成されたビットが 0 の場合

$$b_n \leftarrow 0$$
$$u_{min}[n] \leftarrow u_{min}[n-1]$$
$$u_{max}[n] \leftarrow \frac{u_{min}[n-1] + u_{max}[n-1]}{2}$$

- 生成されたビットが 1 の場合

$$b_n \leftarrow 1$$
$$u_{min}[n] \leftarrow \frac{u_{min}[n-1] + u_{max}[n-1]}{2}$$
$$u_{max}[n] \leftarrow u_{max}[n-1]$$

20: (r の範囲を更新)

$$r_{min} \leftarrow F^{-1}(u_{min}[n])$$
$$r_{max} \leftarrow F^{-1}(u_{max}[n])$$

30: (収束判定)
r_{min} と r_{max} が

- 異符号 (0 と非 0 のときを含む) の場合
  $\Rightarrow$ 31 に進む
- 同符号 (どちらも正数である) の場合
  $\Rightarrow$ 32 に進む
- 同符号 (どちらも負数である) の場合
  $\Rightarrow$ 33 に進む

31: (異符号)

- $r_{max} - r_{min} \leq h$ の場合
  RBG によりランダムに 1 ビット生成し (この時生成したビットは $\mathbf{b}$ へ格納しない)、

– 生成されたビットが 0 の場合
⇒ r_{min} を返す。
– 生成されたビットが 1 の場合
⇒ r_{max} を返す。

- $r_{max} - r_{min} > h$ の場合
⇒ 10 に戻る

32: (同符号:正)

- $r_{min} < b$ かつ $r_{max} - r_{min} \leq h$ の場合
- $b \leq r_{min}$ かつ $r_{max} - r_{min} \leq \frac{h}{b} r_{min}$ の場合
⇒ r_{max} を返す
- それ以外の場合
⇒ 10 に戻る

33: (同符号:負)

- $-r_{max} < b$ かつ $r_{max} - r_{min} \leq h$ の場合
- $b \leq -r_{max}$ かつ $r_{max} - r_{min} \leq -\frac{h}{b} r_{max}$ の場合
⇒ r_{min} を返す
- それ以外の場合
⇒ 10 に戻る

以上が、RDG の擬似コードである。

2.3.2.4 解説

以下、順次擬似コードの意味を説明していく。

00: (初期化)
ここでの処理は、最初に $[0, 1]$ 上の一様乱数 u を生成することに対応している。

10: (u の範囲を二等分)
ここでの処理は、u の小数点以下 n ビット目を調べることに対応している。なお、u は一様乱数であったので、0 である確率と 1 である確率は等しい。よって、RBG が 0/1 を生成することとビットが 0/1 であることが対応する。

20: (r の範囲を更新)
ビットを調べることで u の範囲が狭まったので、対応する r の範囲を更新する。

30: (収束判定)
ここで、r の範囲が十分狭くなっているか判定する。

2.4 まとめ

提案手法に関しては以上である。次の章では、この論文の最も大きな目標である、精度が保証されていることの証明を実数にて行う。そのために、

(1) RDG が停止する
(2) RDG が生成する乱数が (b, h) 離散化条件を満たす
(3) RDG が確率密度関数 f に従う乱数を生成している

を順番に証明する。特に (2) により、任意の箇所の近傍に生成可能な乱数が存在することが分かり、(3) と合わせると、テールの精度が (実は全ての領域の精度が) 保証されていることが分かる。

なお、擬似コードにて使用されている、現在の反復回数 n や RBG の出力列 $\mathbf{b}$ は、必ずしもアルゴリズムに必要なものではない。しかしながら、これら一見無駄に思える変数が、証明を平易にするのである。

3. 正当性

3.1 停止性

2.3.2.3 章の擬似コードは、RBG が 0 または 1 の一方のみを出力する時を除き、停止する。

これを証明するために、まずは補題を一つ示す。

3.1.1 補題 1

n 反復目の $u_{min}[n]$ と $u_{max}[n]$ に関して、以下が成立する。

$$u_{min}[n] = \sum_{i=0}^{n} b_i \times \left(\frac{1}{2}\right)^i$$
$$u_{max}[n] = u_{min}[n] + \left(\frac{1}{2}\right)^n$$

3.1.1.1 補題 1 の証明

(i) $n = 0$ のとき
初期化処理より、$b_0 = 0$ であるので、

$$\sum_{i=0}^{0} b_0 \times \left(\frac{1}{2}\right)^i = 0 = u_{min}[0]$$
$$u_{min}[0] + \left(\frac{1}{2}\right)^0 = 1 = u_{max}[0]$$

となり、成立する。

(ii) $n = k$ で成立しているとき
$n = k + 1$ のときを考える。$(k + 1)$ 反復目において、

(a) 擬似コードの 10 にて RBG が 0 を出力した場合
$b_{k+1} = 0$ であるので、仮定とあわせて

$$\begin{aligned} u_{min}[k+1] &= u_{min}[k] \\ &= u_{min}[k] + b_{k+1} \times \left(\frac{1}{2}\right)^{k+1} \\ &= \sum_{i=0}^{k} b_i \times \left(\frac{1}{2}\right)^i + b_{k+1} \times \left(\frac{1}{2}\right)^{k+1} \\ &= \sum_{i=0}^{k+1} b_i \times \left(\frac{1}{2}\right)^i \end{aligned}$$

$$\begin{aligned} u_{max}[k+1] &= \frac{u_{min}[k] + u_{max}[k]}{2} \\ &= \frac{u_{min}[k] + u_{min}[k] + \left(\frac{1}{2}\right)^k}{2} \\ &= u_{min}[k] + \left(\frac{1}{2}\right)^{k+1} \\ &= u_{min}[k+1] + \left(\frac{1}{2}\right)^{k+1} \end{aligned}$$

となり、成立する。

(b) 擬似コードの 10 にて RBG が 1 を出力した場合

$b_{k+1} = 1$ であるので、仮定とあわせて

$$\begin{aligned} u_{min}[k+1] &= \frac{u_{min}[k] + u_{max}[k]}{2} \\ &= \frac{u_{min}[k] + u_{min}[k] + \left(\frac{1}{2}\right)^k}{2} \\ &= u_{min}[k] + b_{k+1} \times \left(\frac{1}{2}\right)^{k+1} \\ &= \sum_{i=0}^{k} b_i \times \left(\frac{1}{2}\right)^i + b_{k+1} \times \left(\frac{1}{2}\right)^{k+1} \\ &= \sum_{i=0}^{k+1} b_i \times \left(\frac{1}{2}\right)^i \end{aligned}$$

$$\begin{aligned} u_{max}[k+1] &= u_{max}[k] \\ &= u_{min}[k] + \left(\frac{1}{2}\right)^k \\ &= u_{min}[k] + \left(\frac{1}{2}\right)^{k+1} + \left(\frac{1}{2}\right)^{k+1} \\ &= \frac{u_{min}[k] + u_{min}[k] + \left(\frac{1}{2}\right)^k}{2} + \left(\frac{1}{2}\right)^{k+1} \\ &= \frac{u_{min}[k] + u_{max}[k]}{2} + \left(\frac{1}{2}\right)^{k+1} \\ &= u_{min}[k+1] + \left(\frac{1}{2}\right)^{k+1} \end{aligned}$$

となり、成立する。

(i)、(ii) より、数学的帰納法から、題意は示せた。

3.1.2 停止性の証明

補題 1 より、RBG が 0 のみを出力する時は、

$$\begin{aligned} r_{min} &= F^{-1}\left(u_{min}[n]\right) \\ &= F^{-1}\left(\sum_{k=0}^{n} 0 \times \left(\frac{1}{2}\right)^k\right) \\ &= F^{-1}\left(0\right) \\ &= -\infty \end{aligned}$$

となり、逆に 1 のみを出力する時は、

$$\begin{aligned} r_{max} &= F^{-1}\left(u_{max}[n]\right) \\ &= F^{-1}\left(u_{min}[n] + \left(\frac{1}{2}\right)^n\right) \\ &= F^{-1}\left(\sum_{k=0}^{n} 1 \times \left(\frac{1}{2}\right)^k + \left(\frac{1}{2}\right)^n\right) \\ &= F^{-1}\left(1\right) \\ &= \infty \end{aligned}$$

となるので、明らかに停止しない。

それ以外の場合に停止することは、累積分布関数 F が次の条件を満たすことを言えば良い。

$$\forall u \in (0,1),\ \exists(\text{finite}) n \in \mathbb{N}$$
$$\text{s.t. } F^{-1}\left(u + \left(\frac{1}{2}\right)^n\right) - F^{-1}\left(u\right) \leq h$$

なぜならば、この u として $u_{min}[n]$ を用いれば、補題 1 より、

$$\begin{aligned} r_{max} - r_{min} &= F^{-1}\left(u_{max}[n]\right) - F^{-1}\left(u_{min}[n]\right) \\ &= F^{-1}\left(u_{min}[n] + \left(\frac{1}{2}\right)^n\right) - F^{-1}\left(u_{min}[n]\right) \\ &\leq h \end{aligned}$$

とできるが、$r_{max} - r_{min} \leq h$ ならば、擬似コードの 31、32、33 のいずれにおいても停止するためである。

なお、実際の証明であるが、ここでは背理法を用いる。すなわち、ある $u \in (0,1)$ が存在して、任意の $n \in \mathbb{N}$ において、

$$F^{-1}\left(u + \left(\frac{1}{2}\right)^n\right) - F^{-1}\left(u\right) > h$$

となると仮定し、矛盾を導く。

F 及び F^{-1} の単調増加性より、

$$\begin{aligned} 0 &= u - u \\ &= F\left(F^{-1}\left(u\right)\right) - u \\ &< F\left(F^{-1}\left(u\right) + h\right) - u \\ &< F\left(F^{-1}\left(u\right) + h\right) \\ &\leq 1 \end{aligned}$$

となるので、

$$\log_{\frac{1}{2}}\left\{F\left(F^{-1}\left(u\right) + h\right) - u\right\}$$

が定義されている。また、

$$n \geq \log_{\frac{1}{2}}\left\{F\left(F^{-1}\left(u\right) + h\right) - u\right\} > 0$$

を満たす n に対して、

$$\left(\frac{1}{2}\right)^n \leq F\left(F^{-1}\left(u\right) + h\right) - u$$

となる。よって、

$$\begin{aligned} F^{-1}\left(u + \left(\frac{1}{2}\right)^n\right) - F^{-1}\left(u\right) &\leq F^{-1}\left(F\left(F^{-1}\left(u\right) + h\right)\right) - F^{-1}\left(u\right) \\ &= F^{-1}\left(u\right) + h - F^{-1}\left(u\right) \\ &= h \end{aligned}$$

となり、仮定に矛盾する。

以上、背理法により題意は示せた。

3.2 出現値間隔

(b,h) 間隔条件と (b,h) 離散化条件をいずれも満たす。

3.2.1 (b,h) 間隔条件を満たすことの証明

RDG が x を出力して停止したとき、停止時の

$$\begin{cases} n \\ \mathbf{b} = b_0\,(=0)\,b_1b_2\cdots b_n \\ u_{min}[n] \\ u_{max}[n] \\ r_{min} = F^{-1}\left(u_{min}[n]\right) \\ r_{max} = F^{-1}\left(u_{max}[n]\right) \end{cases}$$

をそれぞれ、

$$\begin{cases} N \\ \mathbf{B} = B_0\,(=0)\,B_1B_2\cdots B_n \\ U_{MIN} \\ U_{MAX} \\ R_{MIN} = F^{-1}\left(U_{MIN}\right) \\ R_{MAX} = F^{-1}\left(U_{MAX}\right) \end{cases}$$

とおくと、補題 1 より、

$$U_{MIN} = \sum_{k=0}^{N} B_k \times \left(\frac{1}{2}\right)^k$$

$$U_{MAX} = U_{MIN} + \left(\frac{1}{2}\right)^N$$

となっている。

ここで、$x<0$、$x>0$、$x=0$ それぞれの場合について、x と (b,h) 距離条件を満たすような x の右側と左側の元の存在を示していく。

なお、(b,h) 距離条件を満たすとは、二つの値 p、q に対して、以下を満たすことを言う。(こちらに関しても $0<h<<b$ とする。)

(b,h) 距離条件

p と q のうち絶対値の小さい方を m とおく。

- $|m|<b$ のとき
  $|p-q|\le h$ が成立する。
- $|m|\ge b$ のとき
  $|p-q|\le \frac{h}{b}|m|$ が成立する。

自明ではあるが、

$$|m|\ge b \text{ かつ } |p-q|\le h \Rightarrow |p-q|\le h\le \frac{h}{b}|m|$$

をしばしば利用している。(もし $|p-q|\le h$ ならば、p と q は (b,h) 距離条件を満たす。) また、r_{min} と r_{max} が (b,h) 距離条件を満たすことと上記の擬似コードが停止することが同値であることもしばしば利用している。

(i) $x<0$ の場合

擬似コードより、負の数を出力しうるのは 31 と 33 のときのみである。33 では R_{MIN} を出力しており、また、31 においても $R_{MIN}\le 0$、$R_{MAX}\ge 0$ となるため、いずれにせよ R_{MIN} を出力している。よって、$R_{MIN}=x<0$ である。

(a) x が 31 で出力されていた場合

(r) 右側の元の存在

同一の $\mathbf{B}$ に対して、31 で RBG が 1 を返したとき、RDG は R_{MAX} を返す。また、仮定より、R_{MIN} と R_{MAX} は (b,h) 距離条件を満たす。よって、$x=R_{MIN}$ の右側に R_{MAX} という元が存在する。

(l) 左側の元の存在

33 にて、$r_{max}=x=R_{MIN}$ を満たしつつ r_{min} を出力して停止するような RBG の出力列が存在すれば、そのとき出力される r_{min} と x は、(b,h) 距離条件を満たす。よって、そのような RBG の出力列を少なくとも一つ示せば良い。

さて、$B_1,B_2,\cdots,B_N$ のうち少なくとも 1 つは 1 である。なぜならば、もし $B_1,B_2,\cdots,B_N$ が全て 0 であると仮定すると、補題 1 より $U_{MIN}=0$ となるが、このとき、$R_{MIN}=F^{-1}(0)=-\infty$ となり、擬似コードのアルゴリズムは停止していなかったはずであり、矛盾する。よって、$B_1,B_2,\cdots,B_N$ のうち少なくとも 1 つは 1 が存在する。そこで、$B_1,B_2,\cdots,B_N$ のうち最後の 1 を B_M とおく。$(M\le N)$

また、$\mathbf{C}$ を次のようにとる。

$$\mathbf{C} = B_0B_1B_2\cdots B_{M-1}0111\cdots \text{以降全て } 1$$

このような $\mathbf{C}$ に対して、擬似コードの 10 にて RBG が i 回目に出力するビットが C_i であるときを考える。このとき、$n<M$ となる n で反復が停止することは無い。なぜならば、$\mathbf{C}$ の定義から、$\mathbf{C}$ は最初の M ビットが $\mathbf{B}$ と同一であり、もし $n<M$ なる n で反復が停止するならば、RDG は x を出力する前に停止していたはずであり、RDG が x を出力する際の反復回数 N と矛盾する。

さて、$n\,(\ge M)$ 回目の反復において、$\mathbf{C}$ の定義と補題 1 より、

$$\begin{aligned} u_{min}[n] &= \sum_{k=0}^{M-1} C_k \times \left(\frac{1}{2}\right)^k + \sum_{k=M}^{n} C_k \times \left(\frac{1}{2}\right)^k \\ &= \sum_{k=0}^{M} B_k \times \left(\frac{1}{2}\right)^k - \left(\frac{1}{2}\right)^M \\ &\quad + \sum_{k=M}^{n} 1 \times \left(\frac{1}{2}\right)^k - \left(\frac{1}{2}\right)^M \\ &= \sum_{k=0}^{N} B_k \times \left(\frac{1}{2}\right)^k - \left(\frac{1}{2}\right)^n \\ &= U_{MIN} - \left(\frac{1}{2}\right)^n \end{aligned}$$

$$u_{max}[n] = u_{min}[n] + \left(\frac{1}{2}\right)^n$$
$$= U_{MIN}$$

となる。
また、このときの (擬似コードの 10 にて RBG が i 回目に出力するビットが C_i であるような)RDG が有限回の反復で停止することは停止性より明らかであり、停止時においては、

$$r_{max} = F^{-1}\left(u_{max}[n]\right) = R_{MIN} < 0$$
$$r_{min} < r_{max} < 0$$

より、確かに 33 で r_{min} を出力して停止している。よって、x の左側の元の存在は示せた。

(b) x が 33 で出力されていた場合

擬似コードより、R_{MIN} と R_{MAX} は同符号である。よって、$R_{MAX} < 0$ である。

(r) 右側の元の存在

仮定より、$R_{MIN}\,(= x)$ と R_{MAX} は (b, h) 距離条件を満たす。よって、31 または 33 にて R_{MAX} を出力して停止するような RBG の出力列を少なくとも一つ示せば良い。
さて、$B_1, B_2, \cdots, B_N$ のうち少なくとも 1 つは 0 である。なぜならば、もし $B_1, B_2, \cdots, B_N$ が全て 1 であると仮定すると、補題 1 より $U_{MAX} = 1$ となるが、このとき、$R_{MAX} = F^{-1}(1) = +\infty$ となり、擬似コードのアルゴリズムは停止していなかったということになり、矛盾する。よって、$B_1, B_2, \cdots, B_N$ のうち少なくとも 1 つは 0 が存在する。そこで、$B_1, B_2, \cdots, B_N$ のうち最後の 0 を B_M とおく。$(M \leq N)$
また、$\mathbf{C}$ を次のようにとる。

$$\mathbf{C} = B_0 B_1 B_2 \cdots B_{M-1} 1000 \cdots \text{以降全て } 0$$

このような $\mathbf{C}$ に対して、擬似コードの 10 にて RBG が i 回目に出力するビットが C_i であるときを考える。このとき、$n < M$ となる n で反復が停止することは無い。なぜならば、$\mathbf{C}$ の定義から、$\mathbf{C}$ は最初の M ビットが $\mathbf{B}$ と同一であり、もし $n < M$ なる n で反復が停止するならば、RDG は x を出力する前に停止していたはずであり、RDG が x を出力する際の反復回数 N と矛盾する。
さて、$n\,(\geq M)$ 回目の反復において、$\mathbf{C}$ の定義と補題 1 より、

$$\begin{aligned}
u_{min}[n] &= \sum_{k=0}^{n} C_k \times \left(\frac{1}{2}\right)^k \\
&= \sum_{k=0}^{M} C_k \times \left(\frac{1}{2}\right)^k \\
&= \sum_{k=0}^{M-1} C_k \times \left(\frac{1}{2}\right)^k + \left(\frac{1}{2}\right)^M \\
&= \sum_{k=0}^{M-1} B_k \times \left(\frac{1}{2}\right)^k + \left(\frac{1}{2}\right)^M \\
&= \sum_{k=0}^{N} B_k \times \left(\frac{1}{2}\right)^k - \sum_{k=M}^{N} B_k \times \left(\frac{1}{2}\right) + \left(\frac{1}{2}\right)^M \\
&= \sum_{k=0}^{N} B_k \times \left(\frac{1}{2}\right)^k - \left\{\sum_{k=M}^{N} 1 \times \left(\frac{1}{2}\right) - \left(\frac{1}{2}\right)^M\right\} \\
&\quad + \left(\frac{1}{2}\right)^M \\
&= \sum_{k=0}^{N} B_k \times \left(\frac{1}{2}\right)^k + \left(\frac{1}{2}\right)^{M-1} \\
&= U_{MIN} + \left(\frac{1}{2}\right)^N \\
&= U_{MAX}
\end{aligned}$$

$$\begin{aligned}
u_{max}[n] &= u_{min}[n] + \left(\frac{1}{2}\right)^n \\
&= U_{MAX} + \left(\frac{1}{2}\right)^n
\end{aligned}$$

となる。
また、このときの (擬似コードの 10 にて RBG が i 回目に出力するビットが C_i であるような)RDG が有限回の反復で停止することは停止性より明らかであり、停止時においては、

$$\begin{aligned}
r_{min} &= F^{-1}\left(u_{min}[n]\right) \\
&= F^{-1}\left(U_{MAX}\right) \\
&= R_{MAX} < 0
\end{aligned}$$

となるので、$r_{max} < 0$ ならば 33 にて $r_{min} = R_{MAX}$ を、$r_{max} \geq 0$ ならば 31 にて (RBG が 0 を出力したときに)$r_{min} = R_{MAX}$ を出力して停止している。よって、$x = R_{MIN}$ の右側に R_{MAX} という元が存在する。

(l) 左側の元の存在

(i)-(a)-(l) と同様である。

(ii) $x > 0$ の場合

擬似コードより、正の数を出力しうるのは 31 と 32 のときのみである。32 では R_{MAX} を出力しており、また、31 においても $R_{MIN} \leq 0$、$R_{MAX} \geq 0$ となるため、いずれにせよ R_{MAX} を出力している。つまり、$R_{MAX} = x > 0$ である。

(c) x が 31 で出力されていた場合

(r) 右側の元の存在

32 にて、$r_{min} = x = R_{MAX}$ を満たしつつ r_{max} を出力して停止するような RBG の出力列が存在すれば、そのとき出力される r_{max} と x は、(b, h) 距離条件を満たす。よって、そのような RBG の出力列を少なくとも一つ示せば良い。

ここで (i)-(b)-(r) と同様に、$B_1, B_2, \cdots, B_N$ のうち最後の 0 を B_M とおき $(M \leq N)$、$\mathbf{C}$ を次のようにとる。

$$\mathbf{C} = B_0 B_1 B_2 \cdots B_{M-1} 1000 \cdots \text{以降全て } 0$$

このような $\mathbf{C}$ に対して、擬似コードの 10 にて RBG が i 回目に出力するビットが C_i であるときを考える。このとき、$n < M$ となる n で反復が停止しないこと、及び、有限回の反復で停止することは、(i)-(b)-(r) で示されている。

さて、停止時においては、(i)-(b)-(r) と同様の導出により、

$$\begin{aligned} u_{min}[n] &= U_{MAX} \\ u_{max}[n] &= U_{MAX} + \left(\frac{1}{2}\right)^n \\ r_{min} &= F^{-1}\left(u_{min}[n]\right) \\ &= R_{MAX} \\ &> 0 \\ r_{max} &> r_{min} \\ &> 0 \end{aligned}$$

となるので、確かに 32 で r_{max} を出力して停止している。よって、x の右側の元の存在は示せた。

(l) 左側の元の存在

同一の $\mathbf{B}$ に対して、31 で RBG が 0 を返したとき、RDG は R_{MIN} を返す。また、仮定より、R_{MIN} と R_{MAX} は (b, h) 距離条件を満たす。よって、$x = R_{MAX}$ の左側に R_{MIN} という元が存在する。

(d) x が 32 で出力されていた場合

擬似コードより、R_{MIN} と R_{MAX} は同符号である。よって、$R_{MIN} > 0$ である。

(r) 右側の元の存在

(ii)-(c)-(r) と同様である。

(l) 左側の元の存在

31 または 32 にて、$r_{max} = x = R_{MIN}$ を満たしつつ r_{min} を出力して停止するような RBG の出力列が存在すれば、そのとき出力される r_{min} と x は、(b, h) 距離条件を満たす。よって、そのような RBG の出力列を少なくとも一つ示せば良い。

ここで (i)-(a)-(l) と同様に、$B_1, B_2, \cdots, B_N$ のうち最後の 1 を B_M とおき $(M \leq N)$、$\mathbf{C}$ を次のようにとる。

$$\mathbf{C} = B_0 B_1 B_2 \cdots B_{M-1} 0111 \cdots \text{以降全て } 1$$

このような $\mathbf{C}$ に対して、擬似コードの 10 にて RBG が i 回目に出力するビットが C_i であるときを考える。このとき、$n < M$ となる n で反復が停止しないこと、及び、有限回の反復で停止することは、(i)-(a)-(l) で示されている。

さて、停止時においては、(i)-(a)-(l) と同様の導出により、

$$\begin{aligned} u_{min}[n] &= U_{MIN} - \left(\frac{1}{2}\right)^n \\ u_{max}[n] &= U_{MIN} \\ r_{max} &= F^{-1}\left(u_{max}[n]\right) \\ &= F^{-1}\left(U_{MIN}\right) \\ &= R_{MIN} \\ &> 0 \end{aligned}$$

となるので、$r_{min} > 0$ ならば 32 にて $r_{max} = R_{MIN}$ を、$r_{min} \leq 0$ ならば 31 にて (RBG が 1 を出力したときに)$r_{max} = R_{MIN}$ を出力して停止している。よって、$x = R_{MAX}$ の左側に R_{MIN} という元が存在する。

(iii) $x = 0$ の場合

擬似コードより、0 を出力し得るのは 31 のときのみである。また、$R_{MIN} < R_{MAX}$ となるので、R_{MIN} と R_{MAX} のうち、0 となるのはどちらか一方のみである。

(e) $R_{MIN} = 0$ の場合

$R_{MIN} < R_{MAX}$ より、$R_{MAX} > 0$ である。

(r) 右側の元の存在

(i)-(a)-(r) と同様である。

(l) 左側の元の存在

(i)-(a)-(l) と同様である。ただし、$R_{MIN} = 0$ であるので、擬似コードの 33 ではなく 31 にて (RBG が 0 を出力したときに)r_{min} を出力して停止している。

(f) $R_{MAX} = 0$ の場合

$R_{MIN} < R_{MAX}$ より、$R_{MIN} < 0$ である。

(r) 右側の元の存在

(ii)-(c)-(r) と同様である。ただし、$R_{MAX} = 0$ であるので、擬似コードの 32 ではなく 31 にて (RBG が 1 を出力したときに)r_{max} を出力して停止している。

(l) 左側の元の存在

(ii)-(c)-(l) と同様である。

3.2.2 (b,h) 離散化条件を満たすことの証明

任意の $r \in \mathbb{R}$ に対して、$\mathbb{DR}_f$ 上において、r と (b,h) 距離条件を満たすような r の右側と左側の元の存在を示せば良い。

擬似コードより、

$$u_{min}[0] = 0 \leq F(r) \leq 1 = u_{max}[0]$$

及び、

$$\forall k \in \mathbb{N},\ (u_{min}[k-1] \leq F(r) \leq u_{max}[k-1] \Rightarrow \exists b_k \in \{0,1\}\ \text{s.t.}\ u_{min}[k] \leq F(r) \leq u_{max}[k])$$

が成立するので、

$$u_{min}[n] \leq F(r) \leq u_{max}[n]$$

を (停止するまで) 常に満たし続けるような **b**(擬似コードの 10 における RBG の出力列) が存在する。
また、停止性より、このときの RDG は、

$$b_1, b_2, \cdots, b_n$$

が全て 0 または全て 1 のときを除いて停止するが、

- 全て 0 で停止しないと仮定すると
  すなわち、

  $$\forall n \in \mathbb{N},\ u_{min}[n] \leq F(r) \leq u_{max}[n]$$

  であるので、補題 1 より、

  $$\left(\forall n \in \mathbb{N},\ 0 \leq F(r) \leq \left(\frac{1}{2}\right)^n\right) \Rightarrow F(r) = 0 \Rightarrow r = -\infty$$

  となる。
- 全て 1 で停止しないと仮定すると
  すなわち、

  $$\forall n \in \mathbb{N},\ u_{min}[n] \leq F(r) \leq u_{max}[n]$$

  であるので、補題 1 より、

  $$\left(\forall n \in \mathbb{N},\ 1 - \left(\frac{1}{2}\right)^n \leq F(r) \leq 1\right) \Rightarrow F(r) = 1 \Rightarrow r = +\infty$$

  となる。

ゆえに、結局、任意の $r \in \mathbb{R}$ で停止することが分かる。
さて、停止時においては、

$$\begin{aligned} r_{min} &= F^{-1}(u_{min}[n]) \\ &\leq F^{-1}(F(r)) = r \\ &\leq F^{-1}(u_{max}[n]) = r_{max} \end{aligned}$$

となっており、かつ、(b,h) 間隔条件の証明において、この r_{min} と r_{max} はいずれも $\mathbb{DR}_f$ の元であることが示されている。ここで、$r = r_{min}$ または $r = r_{max}$ であるときは $r \in \mathbb{DR}_f$ となるため、(b,h) 間隔条件の議論と同様になるので省略し、$r \neq r_{min}$ かつ $r \neq r_{max}$ であるときのみを考える。

(i) r_{min} と r_{max} が異符号 (0 と非 0 の場合を含む) の時
擬似コードより、

$$|r_{max} - r_{min}| < h$$

となる。また、

$$|r - r_{min}| < |r_{max} - r_{min}|$$
$$|r_{max} - r| < |r_{max} - r_{min}|$$

であるので、r は r_{min} と r_{max} のそれぞれと (b,h) 距離条件を満たしている。

(ii) r_{min} と r_{max} がともに正であるとき

$$0 < |r_{min}| < |r| < |r_{max}|$$

であるので、

$$\frac{h}{b}|r_{min}| < \frac{h}{b}|r|$$

となる。また

$$|r - r_{min}| < |r_{max} - r_{min}|$$
$$|r_{max} - r| < |r_{max} - r_{min}|$$

となるので、r は r_{min} と r_{max} のそれぞれと (b,h) 距離条件を満たしている。

(iii) r_{min} と r_{max} がともに負であるとき

$$0 < |r_{max}| < |r| < |r_{min}|$$

より、

$$\frac{h}{b}|r_{max}| < \frac{h}{b}|r|$$

となるので、(ii) と同様に、r は r_{min} と r_{max} のそれぞれと (b,h) 距離条件を満たしている。

よって、r の左右に r_{min} と r_{max} という元が存在し、それぞれ r と (b,h) 距離条件を満たしている。

3.3 乱数生成確率

確率密度関数 f に従っている。つまり、丸め健全な $round_{\mathbb{DR}_f}$ が少なくとも一つ存在し、任意の $d \in \mathbb{DR}_f$ に対して、RDG が d を出力する確率 $P(d)$ が、

$$P(d) = \int_{round^{-1}_{\mathbb{DR}_f}[\{d\}]} f(r)\,dr$$

となる。

これを証明するために、まずはいくつかの補題を示す。

3.3.1 補題 2

RDG が x を出力して停止したとき、停止時の

$$\begin{cases} n \\ \mathbf{b} = b_0\,(=0)\,b_1 b_2 \cdots b_n \\ u_{min}[n] \\ u_{max}[n] \\ r_{min} = F^{-1}\left(u_{min}[n]\right) \\ r_{max} = F^{-1}\left(u_{max}[n]\right) \end{cases}$$

をそれぞれ、

$$\begin{cases} N \\ \mathbf{B} = B_0\,(=0)\,B_1 B_2 \cdots B_n \\ U_{MIN} \\ U_{MAX} \\ R_{MIN} = F^{-1}\left(U_{MIN}\right) \\ R_{MAX} = F^{-1}\left(U_{MAX}\right) \end{cases}$$

とおくと、出現値集合 $\mathbb{DR}_f$ に関して、

$$\begin{cases} R_{MIN} \in \mathbb{DR}_f \\ R_{MAX} \in \mathbb{DR}_f \end{cases} \tag{2}$$

及び、

$$\forall r \in (R_{MIN}, R_{MAX})\,,\, r \notin \mathbb{DR}_f \tag{3}$$

が成立する。

3.3.1.1 補題 2 の証明

(2) は 3.2.1 章にて既に示されているため省略し、(3) に関して背理法により証明する。

$\mathbb{DR}_f$ の元 r で、$R_{MIN} < r < R_{MAX}$ となるものが存在する、すなわち、RDG は r を出力しうると仮定し、RDG が r を出力して停止したとき、停止時の

$$\begin{cases} n \\ \mathbf{b} = b_0\,(=0)\,b_1 b_2 \cdots b_n \\ u_{min}[n] \\ u_{max}[n] \\ r_{min} = F^{-1}\left(u_{min}[n]\right) \\ r_{max} = F^{-1}\left(u_{max}[n]\right) \end{cases}$$

をそれぞれ、

$$\begin{cases} n' \\ \mathbf{b}' = b'_0\,(=0)\,b'_1 b'_2 \cdots b'_n \\ u'_{min} \\ u'_{max} \\ r'_{min} = F^{-1}\left(u'_{min}\right) \\ r'_{max} = F^{-1}\left(u'_{max}\right) \end{cases}$$

とおく。

さて、擬似コードより、RDG は r として r'_{min} または r'_{max} のいずれかを出力しているはずなので、

$$R_{MIN} < r'_{min} < R_{MAX}$$
$$R_{MIN} < r'_{max} < R_{MAX}$$

のうち、少なくとも一方が成立する。各式各項に F を作用させることで、これは、

$$U_{MIN} < u'_{min} < U_{MAX}$$
$$U_{MIN} < u'_{max} < U_{MAX}$$

となる。

また、$r \neq x$ より、$b'_M \neq B_M$ となる $M \leq \min(n', N)$ が存在する。なぜなら、もしそのような M が存在しないならば、$\mathbf{B}$ と $\mathbf{b}'$ は一方が他方の接頭辞となっているはずであり、$\mathbf{B}$ が $\mathbf{b}'$ の接頭辞であるときは、RDG は r を出力する前に $x \neq r$ を出力して停止するはずであり、$\mathbf{b}'$ が $\mathbf{B}$ の接頭辞であるときは、RDG は x を出力する前に $r \neq x$ を出力して指定するはずであるので、いずれの場合も矛盾する。

以下、B_M と b'_M の値により場合分けを行う。$B_M \in \{0,1\}$ かつ $b'_M \in \{0,1\}$ より、

$$(B_M, b'_M) = \begin{cases} (0,1) \\ (1,0) \end{cases}$$

である。

(i) $B_M = 0, b'_M = 1$ のとき

$$\begin{aligned} U_{MAX} &= U_{MIN} + \left(\frac{1}{2}\right)^N \\ &= \sum_{k=0}^{N} B_k \times \left(\frac{1}{2}\right)^k + \left(\frac{1}{2}\right)^N \\ &\leq \sum_{k=0}^{M} B_k \times \left(\frac{1}{2}\right)^k + \left(\frac{1}{2}\right)^M \\ &= \sum_{k=0}^{M-1} B_k \times \left(\frac{1}{2}\right)^k + B_M \times \left(\frac{1}{2}\right)^M + \left(\frac{1}{2}\right)^M \\ &= \sum_{k=0}^{M-1} B_k \times \left(\frac{1}{2}\right)^k + \left(\frac{1}{2}\right)^M \\ &= \sum_{k=0}^{M-1} b'_k \times \left(\frac{1}{2}\right)^k + b'_M \times \left(\frac{1}{2}\right)^M \\ &= \sum_{k=0}^{M} b'_k \times \left(\frac{1}{2}\right)^k \\ &\leq \sum_{k=0}^{n'} b'_k \times \left(\frac{1}{2}\right)^k \\ &= u'_{min} \\ &< u'_{max} \end{aligned}$$

よって、

$$U_{MIN} < u'_{min} < U_{MAX}$$
$$U_{MIN} < u'_{max} < U_{MAX}$$

をいずれも満たさないため、矛盾する。

(ii) $B_M = 1, b'_M = 0$ のとき

$$\begin{aligned}
u'_{min} &< u'_{max} \\
&= u'_{min} + \left(\frac{1}{2}\right)^{n'} \\
&= \sum_{k=0}^{n'} b'_k \times \left(\frac{1}{2}\right)^k + \left(\frac{1}{2}\right)^{n'} \\
&\leq \sum_{k=0}^{M} b'_k \times \left(\frac{1}{2}\right)^k + \left(\frac{1}{2}\right)^{M} \\
&= \sum_{k=0}^{M-1} b'_k \times \left(\frac{1}{2}\right)^k + b'_M \times \left(\frac{1}{2}\right)^{n'} + \left(\frac{1}{2}\right)^{M} \\
&= \sum_{k=0}^{M-1} b'_k \times \left(\frac{1}{2}\right)^k + \left(\frac{1}{2}\right)^{M} \\
&= \sum_{k=0}^{M-1} B_k \times \left(\frac{1}{2}\right)^k + B_M \times \left(\frac{1}{2}\right)^{M} \\
&= \sum_{k=0}^{M} B_k \times \left(\frac{1}{2}\right)^k \\
&\leq \sum_{k=0}^{N} B_k \times \left(\frac{1}{2}\right)^k \\
&= U_{MIN}
\end{aligned}$$

よって、

$$U_{MIN} < u'_{min} < U_{MAX}$$
$$U_{MIN} < u'_{max} < U_{MAX}$$

をいずれも満たさないため、矛盾する。

以上、いずれの場合も矛盾するため、背理法より、題意は示せた。

3.3.2 補題 3

RDG が x を出力して停止するような、停止時の (r_{min}, r_{max}) に対して、以下が成り立つ。

(i) $x < 0$ のとき

停止時の (r_{min}, r_{max}) は一意に定まり、かつ、$x = r_{min}$ を満たす。

(ii) $x = 0$ のとき

停止時の (r_{min}, r_{max}) は、擬似コードの 31 にて RBG が 0 を出力したときと 1 を出力したときのそれぞれにおいて一意に定まり、かつ、前者は $r_{min} = 0$ を満たし、後者は $r_{max} = 0$ を満たす。

(iii) $x > 0$ のとき

停止時の (r_{min}, r_{max}) は一意に定まり、かつ、$x = r_{max}$ を満たす。

3.3.2.1 補題 3 の証明

(i) $x < 0$ のとき

擬似コードより、負の数を出力しうるのは 31 と 33 のときのみである。33 では r_{min} を出力しており、また、31 においても $r_{min} \leq 0$、$r_{max} \geq 0$ となるため、いずれにせよ r_{min} を出力している。よって、$x = r_{min}$ である。

次に停止時の (r_{min}, r_{max}) の一意性であるが、停止時おける (r_{min}, r_{max}) を任意に 2 つとり、それぞれ $(x, R1_{MAX})$、$(x, R2_{MAX})$ とおく。

補題 2 の (2) より、x、$R1_{MAX}$、$R2_{MAX}$ はいずれも $\mathbb{DR}_f$ の元であるはずだが、もしも $R1_{MAX} < R2_{MAX}$ であれば、$x < r < R2_{MAX}$ となる $r = R1_{MAX}$ が $\mathbb{DR}_f$ に含まれていることになり、補題 2 の (3) に矛盾する。一方 $R1_{MAX} > R2_{MAX}$ のときも同様に、$x < r < R1_{MAX}$ となる $r = R2_{MAX}$ が $\mathbb{DR}_f$ に含まれていることになり、矛盾する。よって、$R1_{MAX} = R2_{MAX}$ となる。

しかし、先ほど $(x, R1_{MAX})$ と $(x, R2_{MAX})$ は任意にとっていたので、結局、停止時の (r_{min}, r_{max}) が一意であったということになる。

(ii) $x = 0$ のとき

擬似コードより、0 を出力しうるのは 31 のときのみである。また、31 にて RBG が 0 を出力したとき、RDG は r_{min} を出力しているのであるから、そのとき $r_{min} = x = 0$ となるのは明らかである。同様に、RBG が 1 を出力したとき、RDG は r_{max} を出力しているのであるから、そのとき $r_{max} = x = 0$ となるのも明らかである。

次に、31 にて RBG が 0 を出力したときと、1 を出力したときのそれぞれについて、停止時の (r_{min}, r_{max}) の一意性であるが、

(a) 31 にて RBG が 0 を出力したとき

停止時における (r_{min}, r_{max}) を任意に 2 つとり、それぞれ $(0, R1_{MAX})$、$(0, R2_{MAX})$ とおく。

補題 2 の (2) より、0、$R1_{MAX}$、$R2_{MAX}$ はいずれも $\mathbb{DR}_f$ の元であるはずだが、もしも $R1_{MAX} < R2_{MAX}$ であれば、$0 < r < R2_{MAX}$ となる $r = R1_{MAX}$ が $\mathbb{DR}_f$ に含まれていることになり、補題 2 の (3) に矛盾する。一方 $R1_{MAX} > R2_{MAX}$ のときも同様に、$0 < r < R1_{MAX}$ となる $r = R2_{MAX}$ が $\mathbb{DR}_f$ に含まれていることになり、矛盾する。よって、$R1_{MAX} = R2_{MAX}$ となる。

しかし、先ほど $(0, R1_{MAX})$ と $(0, R2_{MAX})$ は任意にとっていたので、結局、停止時の (r_{min}, r_{max}) が一意であったということになる。

(b) 31 にて RBG が 1 を出力したとき

停止時における (r_{min}, r_{max}) を任意に 2 つとり、それぞれ $(R1_{MIN}, 0)$、$(R2_{MIN}, 0)$ とおく。

補題2の(2)より、$R1_{MIN}$、$R2_{MIN}$、0はいずれも$\mathbb{DR}_f$の元であるはずだが、もしも$R1_{MIN} < R2_{MIN}$であれば、$R1_{MIN} < r < 0$となる$r = R2_{MIN}$が$\mathbb{DR}_f$に含まれていることになり、補題2の(3)に矛盾する。一方$R1_{MIN} > R2_{MIN}$のときも同様に、$R2_{MIN} < r < 0$となる$r = R1_{MIN}$が$\mathbb{DR}_f$に含まれていることになり、矛盾する。よって、$R1_{MIN} = R2_{MIN}$となる。しかし、先ほど$(R1_{MIN}, 0)$と$(R2_{MIN}, 0)$は任意にとっていたので、結局、停止時の(r_{min}, r_{max})が一意であったということになる。

(iii) $x > 0$のとき

擬似コードより、正の数を出力しうるのは31と32のときのみである。32ではr_{max}を出力しており、また、31においても$r_{min} \leq 0$、$r_{max} \geq 0$となるため、いずれにせよr_{max}を出力している。よって、$x = r_{max}$である。

次に停止時の(r_{min}, r_{max})の一意性であるが、停止時おける(r_{min}, r_{max})を任意に2つとり、それぞれ$(R1_{MIN}, x)$、$(R2_{MIN}, x)$とおく。

補題2の(2)より、$R1_{MIN}$、$R2_{MIN}$、xはいずれも$\mathbb{DR}_f$の元であるはずだが、もしも$R1_{MIN} < R2_{MIN}$であれば、$R1_{MIN} < r < x$となる$r = R2_{MIN}$が$\mathbb{DR}_f$に含まれていることになり、補題2の(3)に矛盾する。一方$R1_{MIN} > R2_{MIN}$のときも同様に、$R2_{MIN} < r < x$となる$r = R1_{MIN}$が$\mathbb{DR}_f$に含まれていることになり、矛盾する。よって、$R1_{MIN} = R2_{MIN}$となる。しかし、先ほど$(R1_{MIN}, x)$と$(R2_{MIN}, x)$は任意にとっていたので、結局、停止時の(r_{min}, r_{max})が一意であったということになる。

以上、題意は示せた。

3.3.2.2 補題3と乱数生成確率

RDGがxを出力して停止したとき、停止時のr_{min}とr_{max}が与えられれば、RDGがその(r_{min}, r_{max})に至って停止する確率を求めることが可能である。

まず、r_{min}とr_{max}が与えられているので、r_{min}とr_{max}の定義から、

$$u_{min}[n] = F(r_{min})$$
$$u_{max}[n] = F(r_{max})$$

により、$u_{min}[n]$と$u_{max}[n]$を求めることができる。

また、補題1より、このようにして求められた$u_{min}[n]$と$u_{max}[n]$を用いれば、

$$\begin{cases} n &= \log_{\frac{1}{2}}(u_{max}[n] - u_{min}[n]) \\ &= \log_{\frac{1}{2}}(F(r_{max}) - F(r_{min})) \\ b_0 &= 0 \\ b_k &= u_{min}[n] \text{の二進数表記時における小数点以下} k \text{桁目} \end{cases}$$

により、nと$\mathbf{b}$も求めることが可能である。

さらに、擬似コードの31で出力された値であるかどうかは、r_{min}とr_{max}が異符号(0と非0の場合を含む)であるかどうかで判定することが可能である。(異符号ならば31で出力された値である。)

これらの準備のもとで、

- r_{min}とr_{max}が異符号(0と非0の場合を含む)
  擬似コードの31で生成されているので、求める確率は、
  - 10にてRBGがk反復目でb_kを出力する確率
  - 31にてRBGが0/1を出力し、RDGがr_{min}/r_{max}を出力する確率

  の積で表される。
  RBGは0と1を等確率に出力するので、結局、

$$\left(\frac{1}{2}\right)^n \times \left(\frac{1}{2}\right) = \left(\frac{1}{2}\right)^{n+1} = \left(\frac{1}{2}\right)^{\log_{\frac{1}{2}}(F(r_{max}) - F(r_{min})) + 1} = \frac{F(r_{max}) - F(r_{min})}{2}$$

  である。
- r_{min}とr_{max}が同符号(0と非0の場合は除く)
  擬似コードの32または33で生成されているので、求める確率は、
  - 10にてRBGがk反復目でb_kを出力する確率

  で表される。
  RBGは0と1を等確率に出力するので、結局、

$$\left(\frac{1}{2}\right)^n = \left(\frac{1}{2}\right)^{\log_{\frac{1}{2}}(F(r_{max}) - F(r_{min}))} = F(r_{max}) - F(r_{min})$$

  である。

以上の方法により、RDGがその(r_{min}, r_{max})に至って停止し、xを出力する確率を求めることが可能である。

さて、補題3より、$x = 0$ならば停止時の(r_{min}, r_{max})の組はちょうど二通り、$x \neq 0$ならば停止時の(r_{min}, r_{max})の組はただ一つとなる。すなわち、何らかの方法で全て(1組または2組)の(r_{min}, r_{max})を求めれば、それぞれについて上記の方法で確率を求め、足し合わせることで、RDGがxを出力する確率を求めることが可能となる。

3.3.3 補題4

任意の$r \in \mathbb{R}$に対して、

$$\{d \in \mathbb{DR}_f \mid d < r\}$$

には最大元が存在し、

$$\{d \in \mathbb{DR}_f \mid r < d\}$$

には最小元が存在する。

3.3.3.1 補題 4 の証明

$\mathbb{DR}_f$ は (b,h) 離散化条件を満たすので、$w = \max\left(h, \frac{h}{b}|r|\right)$ とおくと、

$$\mathbb{X} = \{d \in \mathbb{DR}_f \mid r - w \leq d < r\}$$
$$\mathbb{Y} = \{d \in \mathbb{DR}_f \mid r < d \leq r + w\}$$

はいずれも空集合にはならない。よって、$\mathbb{X}$ に最大元が、$\mathbb{Y}$ に最小元が存在することを言えば良い。

また、証明のための準備として、関数

$$g(t) = \frac{F(t+h) - F(t)}{2}$$

の $t \in [r - w - h, r + w]$ における最小値を M とおく。F の単調増加性、及び、F は 0 以上 1 以下であることから、

$$\begin{aligned} 0 &< \frac{F(t+h) - F(t)}{2} \\ &\leq \frac{F(t+h)}{2} \\ &\leq \frac{1}{2} \end{aligned}$$

が得られるので、$0 < M \leq \frac{1}{2}$ である。

なお、実際の証明は背理法により行う。

(i) $\mathbb{X}$ に最大元が存在しないと仮定すると

すなわち、

$$\forall d \in \mathbb{X}, \exists d' \in \mathbb{X} \text{ s.t. } d < d'$$

が成立すると仮定すると、

$$\mathbb{U}' = \{F(d) \mid d \in \mathbb{X}\}$$

で定義される $\mathbb{U}'$ は無限集合となる。よって、

$$u_1 < u_2 \leq u_1 + M$$

を満たす $u_1, u_2 \in \mathbb{U}'$ をとることができる。

このような u_1 と u_2 に対して、d_1 と d_2 をそれぞれ、

$$d_1 = F^{-1}(u_1)$$
$$d_2 = F^{-1}(u_2)$$

とおくと、F^{-1} の単調増加性より、$d_1 < d_2$ であること、また、$\mathbb{U}'$ と $\mathbb{X}$ の定義より、$d_1, d_2 \in \mathbb{X} \subset \mathbb{DR}_f$ であることが言える。

(a) $d_1 < 0$ のとき

RDG が d_1 を出力して停止したときを考える。なお、k 反復目における r_{min} と r_{max} の値をそれぞれ、$r_{min}[k]$ と $r_{max}[k]$ で表す。つまり、

$$r_{min}[k] = F^{-1}(u_{min}[k])$$
$$r_{max}[k] = F^{-1}(u_{max}[k])$$

とする。

まず、停止時 (n 反復目) について考えると、$d_1 < 0$ であることから、補題 3 より、

$$r_{min}[n] = d_1$$

となる。また、$d_2 \in \mathbb{DR}_f$ であることから、補題 2 の (2) より、

$$\begin{aligned} d_1 &= r_{min}[n] \\ &< r_{max}[n] \\ &\leq d_2 \end{aligned}$$

となる。よって、n について、補題 1 から、

$$\begin{aligned} n &= \log_{\frac{1}{2}}(u_{max}[n] - u_{min}[n]) \\ &= \log_{\frac{1}{2}}(F(r_{max}[n]) - F(r_{min}[n])) \\ &= \log_{\frac{1}{2}}(F(r_{max}[n]) - F(d_1)) \\ &\geq \log_{\frac{1}{2}}(F(d_2) - F(d_1)) \\ &= \log_{\frac{1}{2}}(u_2 - u_1) \\ &\geq \log_{\frac{1}{2}} M \end{aligned}$$

となる。

次に、$(n-1)$ 反復目の

$$r_{max}[n-1] - r_{min}[n-1]$$

について考える。擬似コードから、

$$r_{max} - r_{min} \leq h$$

であればその時点で RDG は停止してしまうので、RDG が n 反復目を迎えるためには、少なくとも、

$$r_{max}[n-1] - r_{min}[n-1] > h$$

が必要条件となる。

(1) $b_n = 0$ のとき

$r_{max}[n-1] - r_{min}[n-1]$ を求めると、

$$\begin{aligned} &r_{max}[n-1] - r_{min}[n-1] \\ &\quad = F^{-1}(u_{max}[n-1]) - F^{-1}(u_{min}[n-1]) \\ &\quad = F^{-1}\left(u_{min}[n] + \left(\frac{1}{2}\right)^{n-1}\right) \\ &\qquad - F^{-1}(u_{min}[n]) \\ &\quad = F^{-1}\left(F(r_{min}[n]) + \left(\frac{1}{2}\right)^{n-1}\right) \\ &\qquad - F^{-1}(F(r_{min}[n])) \\ &\quad = F^{-1}\left(F(d_1) + \left(\frac{1}{2}\right)^{n-1}\right) \\ &\qquad - F^{-1}(F(d_1)) \\ &\quad \leq F^{-1}(F(d_1) + 2M) - d_1 \end{aligned}$$

となる。また、$d_1 \in \mathbb{X}$ であるので、

$$r - w \leq d_1 < r$$

となり、M の定義より、

$$\begin{aligned} 2M &\leq 2g\left(d_1\right) \\ &= F\left(d_1 + h\right) - F\left(d_1\right) \end{aligned}$$

となるので、F^{-1} の単調増加性より、

$$\begin{aligned} r_{max}[n-1] - r_{min}[n-1] &\leq F^{-1}\left(F\left(d_1\right) + 2M\right) - d_1 \\ &\leq F^{-1}\left(F\left(d_1 + h\right)\right) - d_1 \\ &= d_1 + h - d_1 \\ &= h \end{aligned}$$

となり、矛盾する。

(2) $b_n = 1$ のとき

$r_{max}[n-1] - r_{min}[n-1]$ を求めると、

$$\begin{aligned} &r_{max}[n-1] - r_{min}[n-1] \\ &\quad = F^{-1}\left(u_{max}[n-1]\right) - F^{-1}\left(u_{min}[n-1]\right) \\ &\quad = F^{-1}\left(u_{max}[n]\right) \\ &\qquad - F^{-1}\left(u_{max}[n] - \left(\frac{1}{2}\right)^{n-1}\right) \\ &\quad \leq F^{-1}\left(u_{max}[n]\right) - F^{-1}\left(u_{max}[n] - 2M\right) \end{aligned}$$

となる。また、$d_1 < r_{max}[n] \leq d_2$ であったので、
$d_1, d_2 \in \mathbb{X}$ と併せて、

$$\begin{aligned} r - w - h &\leq d_1 - h \\ &< r_{max}[n] - h = F^{-1}\left(u_{max}[n]\right) - h \\ &\leq d_2 - h \\ &< r - h \end{aligned}$$

となり、M の定義より、

$$\begin{aligned} 2M &\leq 2g\left(F^{-1}\left(u_{max}[n]\right) - h\right) \\ &= u_{max}[n] - F\left(F^{-1}\left(u_{max}[n] - h\right)\right) \end{aligned}$$

となるので、F^{-1} の単調増加性より、

$$\begin{aligned} &r_{max}[n-1] - r_{min}[n-1] \\ &\quad \leq F^{-1}\left(u_{max}[n]\right) - F^{-1}\left(u_{max}[n] - 2M\right) \\ &\quad \leq F^{-1}\left(u_{max}[n]\right) \\ &\qquad - F^{-1}\left(F\left(F^{-1}\left(u_{max}[n]\right) - h\right)\right) \\ &\quad = F^{-1}\left(u_{max}[n]\right) - F^{-1}\left(u_{max}[n]\right) + h \\ &\quad = h \end{aligned}$$

となり、矛盾する。

(b) $0 \leq d_1$ のとき

RDG が d_2 を出力して停止したときを考える。なお、k 反復目における r_{min} と r_{max} の値をそれぞれ、$r_{min}[k]$ と $r_{max}[k]$ で表す。つまり、

$$\begin{aligned} r_{min}[k] &= F^{-1}\left(u_{min}[k]\right) \\ r_{max}[k] &= F^{-1}\left(u_{max}[k]\right) \end{aligned}$$

とする。

まず、停止時 (n 反復目) について考えると、$0 \leq d_1 < d_2$ であることから、補題 3 より、

$$r_{max}[n] = d_2$$

となる。また、$d_1 \in \mathbb{DR}_f$ であることから、補題 2 の (2) より、

$$\begin{aligned} d_1 &\leq r_{min}[n] \\ &< r_{max}[n] = d_2 \end{aligned}$$

となる。よって、n について、補題 1 から、

$$\begin{aligned} n &= \log_{\frac{1}{2}}\left(u_{max}[n] - u_{min}[n]\right) \\ &= \log_{\frac{1}{2}}\left(F\left(r_{max}[n]\right) - F\left(r_{min}[n]\right)\right) \\ &= \log_{\frac{1}{2}}\left(F\left(d_2\right) - F\left(r_{min}[n]\right)\right) \\ &\geq \log_{\frac{1}{2}}\left(F\left(d_2\right) - F\left(d_1\right)\right) \\ &= \log_{\frac{1}{2}}\left(u_2 - u_1\right) \\ &\geq \log_{\frac{1}{2}} M \end{aligned}$$

となる。

次に、$(n-1)$ 反復目の

$$r_{max}[n-1] - r_{min}[n-1]$$

について考える。(i)-(a) と同様に、RDG が n 反復目を迎えるためには、少なくとも、

$$r_{max}[n-1] - r_{min}[n-1] > h$$

が必要条件となる。

(1) $b_n = 0$ のとき

$r_{max}[n-1] - r_{min}[n-1]$ を求めると、

$$\begin{aligned} &r_{max}[n-1] - r_{min}[n-1] \\ &\quad = F^{-1}\left(u_{max}[n-1]\right) - F^{-1}\left(u_{min}[n-1]\right) \\ &\quad = F^{-1}\left(u_{min}[n] + \left(\frac{1}{2}\right)^{n-1}\right) \\ &\qquad - F^{-1}\left(u_{min}[n]\right) \\ &\quad \leq F^{-1}\left(u_{min}[n] + 2M\right) - F^{-1}\left(u_{min}[n]\right) \end{aligned}$$

となる。また、$d_1 \leq r_{min}[n] < d_2$ であったので、

$d_1, d_2 \in \mathbb{X}$ と併せて、

$$\begin{aligned} r - w &\leq d_1 \\ &\leq r_{min}[n] = F^{-1}\left(u_{min}[n]\right) \\ &< d_2 \\ &< r \end{aligned}$$

となり、M の定義より、

$$\begin{aligned} 2M &\leq 2g\left(F^{-1}\left(u_{min}[n]\right)\right) \\ &= F\left(F^{-1}\left(u_{min}[n]\right) + h\right) - u_{min}[n] \end{aligned}$$

となるので、F^{-1} の単調増加性より、

$$\begin{aligned} r_{max}[n-1] - r_{min}[n-1] &\leq F^{-1}\left(u_{min}[n] + 2M\right) - F^{-1}\left(u_{min}[n]\right) \\ &\leq F^{-1}\left(F\left(F^{-1}\left(u_{min}[n]\right) + h\right)\right) \\ &\quad - F^{-1}\left(u_{min}[n]\right) \\ &= F^{-1}\left(u_{min}[n]\right) + h - F^{-1}\left(u_{min}[n]\right) \\ &= h \end{aligned}$$

となり、矛盾する。

(2) $b_n = 1$ のとき

$r_{max}[n-1] - r_{min}[n-1]$ を求めると、

$$\begin{aligned} r_{max}[n-1] - r_{min}[n-1] &= F^{-1}\left(u_{max}[n-1]\right) \\ &\quad - F^{-1}\left(u_{min}[n-1]\right) \\ &= F^{-1}\left(u_{max}[n]\right) \\ &\quad - F^{-1}\left(u_{max}[n] - \left(\frac{1}{2}\right)^{n-1}\right) \\ &= F^{-1}\left(F\left(r_{max}[n]\right)\right) \\ &\quad - F^{-1}\left(F\left(r_{max}[n]\right) - \left(\frac{1}{2}\right)^{n-1}\right) \\ &= F^{-1}\left(F\left(d_2\right)\right) \\ &\quad - F^{-1}\left(F\left(d_2\right) - \left(\frac{1}{2}\right)^{n-1}\right) \\ &\leq d_2 - F^{-1}\left(F\left(d_2\right) - 2M\right) \end{aligned}$$

となる。また、$d_2 \in \mathbb{X}$ であるので、

$$r - w - h \leq d_2 - h < r - h$$

となり、M の定義より、

$$\begin{aligned} 2M &\leq 2g\left(d_2 - h\right) \\ &= F\left(d_2\right) - F\left(d_2 - h\right) \end{aligned}$$

となるので、F^{-1} の単調増加性より、

$$\begin{aligned} r_{max}[n-1] - r_{min}[n-1] &\leq d_2 - F^{-1}\left(F\left(d_2\right) - 2M\right) \\ &\leq d_2 - F^{-1}\left(F\left(d_2 - h\right)\right) \\ &= d_2 - d_2 + h \\ &= h \end{aligned}$$

となり、矛盾する。

(ii) $\mathbb{Y}$ に最小元が存在しないと仮定すると

すなわち、

$$\forall d \in \mathbb{Y},\ \exists d' \in \mathbb{Y} \text{ s.t. } d' < d$$

が成立すると仮定すると、

$$\mathbb{U}' = \left\{F\left(d\right) \mid d \in \mathbb{Y}\right\}$$

で定義される $\mathbb{U}'$ は無限集合となる。よって、

$$u_2 - M \leq u_1 < u_2$$

を満たす $u_1, u_2 \in \mathbb{U}'$ をとることができる。

これ以降の議論は、(i) と同様である。

3.3.4 補題 5

$\mathbb{DR}_f$ に対して、正の最小要素を m^+、負の最大要素を m^- とおき (そのような m^- と m^+ が存在することは、補題 4 より明らかである)、$round_{\mathbb{DR}_f}$ として次のようなものを考えれば、$round_{\mathbb{DR}_f}$ は丸め健全である。

ただし、

$$mid\left(r_1, r_2\right) = F^{-1}\left(\frac{F\left(r_1\right) + F\left(r_2\right)}{2}\right)$$

とする。

なお、F と F^{-1} はともに順序を保存する関数であるので、

$$r_\alpha \leq r_\beta \Rightarrow \begin{cases} mid\left(r, r_\alpha\right) \leq mid\left(r, r_\beta\right) \\ mid\left(r_\alpha, r\right) \leq mid\left(r_\beta, r\right) \end{cases}$$

となる。さらに、

$$\begin{aligned} mid\left(r, r\right) &= F^{-1}\left(\frac{F\left(r\right) + F\left(r\right)}{2}\right) \\ &= F^{-1}\left(F\left(r\right)\right) \\ &= r \end{aligned}$$

である。

$round_{\mathbb{DR}_f}(r)$

(i) $r < m^-$ のとき
$\{d \in \mathbb{DR}_f \mid d \leq r\}$ の最大元を返す。
(ii) $m^- \leq r \leq m^+$ のとき
(a) $0 \notin \mathbb{DR}_f$ のとき
(1) $m^- \leq r < mid(m^-, m^+)$ のとき
m^- を返す。
(2) $mid(m^-, m^+) \leq r \leq m^+$ のとき
m^+ を返す。
(b) $0 \in \mathbb{DR}_f$ のとき
(1) $m^- \leq r < mid(m^-, 0)$ のとき
m^- を返す。
(2) $mid(m^-, 0) \leq r \leq mid(0, m^+)$ のとき
0 を返す。
(3) $mid(0, m^+) < r \leq m^+$ のとき
m^+ を返す。
(iii) $m^+ < r$ のとき
$\{d \in \mathbb{DR}_f \mid r \leq d\}$ の最小元を返す。

3.3.4.1 補題 5 の証明

丸め健全であるための条件である、全域一意性・同一性・連続性を順次示す。

- 全域一意性
  各 $r \in \mathbb{R}$ に対して、$round_{\mathbb{DR}_f}(r)$ がちょうど一つの $d \in \mathbb{DR}_f$ に対応することを言えば良い。
  (i) $r < m^-$ のとき
  $round_{\mathbb{DR}_f}(r)$ の定義である最大元は、存在するならば一意であることが明らかなので、存在性についてのみ証明すればよいが、

$$\{d \in \mathbb{DR}_f \mid d < r\}$$

  に最大元が存在することは補題 4 にて既に示されているので、

$$\{d \in \mathbb{DR}_f \mid d \leq r\}$$

  に最大元が存在することは明らかである。
  (ii) $m^- \leq r \leq m^+$ のとき
  $round_{\mathbb{DR}_f}$ の定義より明らか。
  (iii) $m^+ < r$ のとき
  $round_{\mathbb{DR}_f}(r)$ の定義である最小元は、存在するならば一意であることが明らかなので、存在性についてのみ証明すればよいが、

$$\{d \in \mathbb{DR}_f \mid r < d\}$$

  に最小元が存在することは補題 4 にて既に示されているので、

$$\{d \in \mathbb{DR}_f \mid r \leq d\}$$

  に最小元が存在することは明らかである。
- 同一性
  各 $d \in \mathbb{DR}_f$ に対して、$round_{\mathbb{DR}_f}(d) = d$ を示せば良い。
  (i) $d < m^-$ のとき
  $round_{\mathbb{DR}_f}$ の定義より、$r = d$ のとき、

$$d \in \{d' \in \mathbb{DR}_f \mid d' \leq r = d\}$$

  となるので、この最大元は d となる、つまり、$round_{\mathbb{DR}_f}(d) = d$ である。
  (ii) $m^- \leq d \leq m^+$ のとき
  (a) $0 \notin \mathbb{DR}_f$ のとき
  d として考えられるのは、m^- と m^+ のみであるが、$round_{\mathbb{DR}_f}$ の定義より、確かに、

$$round_{\mathbb{DR}_f}(m^-) = m^-$$
$$round_{\mathbb{DR}_f}(m^+) = m^+$$

  となる。
  (b) $0 \in \mathbb{DR}_f$ のとき
  d として考えられるのは、m^- と 0 と m^+ のみであるが、$round_{\mathbb{DR}_f}$ の定義より、確かに、

$$round_{\mathbb{DR}_f}(m^-) = m^-$$
$$round_{\mathbb{DR}_f}(0) = 0$$
$$round_{\mathbb{DR}_f}(m^+) = m^+$$

  となる。
  (iii) $m^+ < d$ のとき
  $round_{\mathbb{DR}_f}$ の定義より、$r = d$ のとき、

$$d \in \{d' \in \mathbb{DR}_f \mid d' \geq r = d\}$$

  となるので、この最小元は d となる、つまり、$round_{\mathbb{DR}_f}(d) = d$ である。
- 連続性
  各 $p, q \in \mathbb{R}$ に対して、

$$round_{\mathbb{X}}(p) = round_{\mathbb{X}}(q) = x$$

  となるならば、任意の $t \in [0, 1]$ に対して、

$$round_{\mathbb{X}}(t * p + (1 - t) * q) = x$$

  となることを示せば良い。
  なお、$p = q$ であるときや、$t = 0$ または $t = 1$ のときは自明であるので、以下では $p \neq q$ とし、$t \in (0, 1)$ のときのみを示す。
  (i) $p < m^-$ のとき
  $round_{\mathbb{DR}_f}$ の定義より、

$$round_{\mathbb{DR}_f}(p) < m^-$$

である。また、$m^- \leq q$ ならば、

$$m^- \leq round_{\mathbb{DR}_f}(q)$$

となるので、結局 $q < m^-$ のときのみを考えれば良い。
さて、任意の実数 $t \in (0,1)$ に対して、$(t*p+(1-t)*q)$ は p と q の間の値であることから、

$$t*p+(1-t)*q < \max(p,q) < m^-$$

となるので、

$$round_{\mathbb{DR}_f}(t*p+(1-t)*q)$$

の値は、

$$\{d \in \mathbb{DR}_f \mid d \leq t*p+(1-t)*q\}$$

の最大元となる。
また、仮定より、

$$round_{\mathbb{DR}_f}(p) = round_{\mathbb{DR}_f}(q) = x$$

であるので、

$$\{d \in \mathbb{DR}_f \mid d \leq p\}$$
$$\{d \in \mathbb{DR}_f \mid d \leq q\}$$

の最大元はいずれも x である。よって、p と q の間の値、及び、$\max(p,q)$ は、$\mathbb{DR}_f$ に含まれていないことになる。ゆえに、

$$\{d \in \mathbb{DR}_f \mid d \leq t*p+(1-t)*q\}$$

の最大元は、

$$\{d \in \mathbb{DR}_f \mid d \leq \min(p,q)\}$$

の最大元と一致する。無論、これは x と等しいので、結局、

$$round_{\mathbb{DR}_f}(t*p+(1-t)*q) = x$$

となる。

(ii) $m^- \leq p \leq m^+$ のとき
$round_{\mathbb{DR}_f}$ の定義より、

$$m^- \leq round_{\mathbb{DR}_f}(p) \leq m^+$$

である。また、$q < m^-$ ならば、

$$round_{\mathbb{DR}_f}(q) < m^-$$

となり、$m^+ < q$ ならば、

$$m^+ < round_{\mathbb{DR}_f}(q)$$

となるので、結局 $m^- \leq q \leq m^+$ のときのみを考えれば良い。

(a) $0 \notin \mathbb{DR}_f$ のとき

(1) $m^- \leq p < mid(m^-, m^+)$ のとき
$round_{\mathbb{DR}_f}$ の定義より、

$$round_{\mathbb{DR}_f}(p) = m^-$$

である。また、

$$round_{\mathbb{DR}_f}(q) = round_{\mathbb{DR}_f}(p) = m^-$$

となるためには、

$$m^- \leq q < mid(m^-, m^+)$$

が必要である。
さて、任意の実数 $t \in (0,1)$ に対して、$(t*p+(1-t)*q)$ は p と q の間の値であることから、

$$\begin{aligned} m^- &\leq \min(p,q) \\ &< t*p+(1-t)*q \\ &< \max(p,q) < mid(m^-, m^+) \end{aligned}$$

となっているはずなので、

$$round_{\mathbb{DR}_f}(t*p+(1-t)*q) = m^-$$

となる。

(2) $mid(m^-, m^+) \leq p \leq m^+$ のとき
(ii)-(a)-(1) と同様である。

(b) $0 \in \mathbb{DR}_f$ のとき

(1) $m^- \leq p < mid(m^-, 0)$ のとき

(2) $mid(m^-, 0) \leq p \leq mid(0, m^+)$ のとき

(3) $mid(0, m^+) < p \leq m^+$ のとき
いずれも (ii)-(a)-(1) と同様である。

(iii) $m^+ < p$ のとき
$round_{\mathbb{DR}_f}$ の定義より、

$$m^+ < round_{\mathbb{DR}_f}(q)$$

である。また、$q \leq m^+$ ならば、

$$round_{\mathbb{DR}_f}(q) \leq m^+$$

となるので、結局 $m^+ < q$ のときのみを考えれば良い。
さて、任意の実数 $t \in (0,1)$ に対して、$(t*p+(1-t)*q)$ は p と q の間の値であることから、

$$m^+ < \min(p,q) < t*p+(1-t)*q$$

となるので、

$$round_{\mathbb{DR}_f}(t*p+(1-t)*q)$$

の値は、

$$\{d \in \mathbb{DR}_f \mid t*p+(1-t)*q \le d\}$$

の最小元となる。
また、仮定より、

$$round_{\mathbb{DR}_f}(p) = round_{\mathbb{DR}_f}(q) = x$$

であるので、

$$\{d \in \mathbb{DR}_f \mid p \le d\}$$
$$\{d \in \mathbb{DR}_f \mid q \le d\}$$

の最小元はいずれも x である。よって、p と q の間の値、及び、$\min(p,q)$ は、$\mathbb{DR}_f$ に含まれていないことになる。ゆえに、

$$\{d \in \mathbb{DR}_f \mid t*p+(1-t)*q \le d\}$$

の最小元は、

$$\{d \in \mathbb{DR}_f \mid \max(p,q) \le d\}$$

の最小元と一致する。無論、これは x と等しいので、結局、

$$round_{\mathbb{DR}_f}(t*p+(1-t)*q) = x$$

となる。

以上から、$round_{\mathbb{DR}_f}$ は上記の性質全てを満たすため、丸め健全である。

3.3.5 補題 6

補題 5 で定義した $round_{\mathbb{DR}_f}$ に対して、その逆像 $round^{-1}_{\mathbb{DR}_f}[\{d\}]$ は、次のようになる。
ただし、各 $d \in \mathbb{DR}_f$ に対して、$dprev_{\mathbb{DR}_f}(d)$ と $dnext_{\mathbb{DR}_f}(d)$ をそれぞれ次のように定義する。

$$dprev_{\mathbb{DR}_f}(d) : \{d' \in \mathbb{DR}_f \mid d' < d\} \text{の最大元}$$
$$dnext_{\mathbb{DR}_f}(d) : \{d' \in \mathbb{DR}_f \mid d < d'\} \text{の最小元}$$

このような $dprev_{\mathbb{DR}_f}(d)$ と $dnext_{\mathbb{DR}_f}(d)$ が任意の $d \in \mathbb{DR}_f$ で定義されていることは、補題 4 から明らかである。

(i) $d < m^-$ のとき

$$round^{-1}_{\mathbb{DR}_f}[\{d\}] = \left\{r \in \mathbb{R} \mid d \le r < dnext_{\mathbb{DR}_f}(d)\right\}$$

(ii) $m^- \le d \le m^+$ のとき

(a) $0 \notin \mathbb{DR}_f$ のとき

(1) $d = m^-$ のとき

$$round^{-1}_{\mathbb{DR}_f}[\{d\}] = \left\{r \in \mathbb{R} \mid m^- \le r < mid\left(m^-, m^+\right)\right\}$$

(2) $d = m^+$ のとき

$$round^{-1}_{\mathbb{DR}_f}[\{d\}] = \left\{r \in \mathbb{R} \mid mid\left(m^-, m^+\right) \le r \le m^+\right\}$$

(b) $0 \in \mathbb{DR}_f$ のとき

(1) $d = m^-$ のとき

$$round^{-1}_{\mathbb{DR}_f}[\{d\}] = \left\{r \in \mathbb{R} \mid m^- \le r < mid\left(m^-, 0\right)\right\}$$

(2) $d = 0$ のとき

$$round^{-1}_{\mathbb{DR}_f}[\{d\}] = \left\{r \in \mathbb{R} \mid mid\left(m^-, 0\right) \le r \le mid\left(0, m^+\right)\right\}$$

(3) $d = m^+$ のとき

$$round^{-1}_{\mathbb{DR}_f}[\{d\}] = \left\{r \in \mathbb{R} \mid mid\left(0, m^+\right) < r \le m^+\right\}$$

(iii) $m^+ < d$ のとき

$$round^{-1}_{\mathbb{DR}_f}[\{d\}] = \left\{r \in \mathbb{R} \mid dprev_{\mathbb{DR}_f}(d) < r \le d\right\}$$

3.3.5.1 補題 6 の証明

逆像の定義より、各 $d \in \mathbb{DR}_f$ に対して、$round_{\mathbb{DR}_f}(r) = d$ となる $r \in \mathbb{R}$ の範囲を求めればよい。

(i) $d < m^-$ のとき

$round_{\mathbb{DR}_f}$ の定義より、$d < m^-$ ならば、

$$round_{\mathbb{DR}_f}(d) < m^-$$

であり、$m^- \le r$ ならば、

$$d < m^- \le round_{\mathbb{DR}_f}(r)$$

であるので、

$$round_{\mathbb{DR}_f}(r) = d$$

となるためには、$r < m^-$ が必要条件となる。

(a) $r < d$ のとき

$$\{d' \in \mathbb{DR}_f \mid d' \le r < d\}$$

の最大元は明らかに d 未満であり、$round_{\mathbb{DR}_f}(r) = d$ となりえない。

(b) $d \le r < dnext_{\mathbb{DR}_f}(d)$ のとき

このとき、

$$\left\{d' \in \mathbb{DR}_f \mid d' \le r < dnext_{\mathbb{DR}_f}(d)\right\}$$

は d を含んでおり、かつ、$dnext_{\mathbb{DR}_f}$ の定義から、$\mathbb{DR}_f$ は d と $dnext_{\mathbb{DR}_f}(d)$ の間に元を持たないので、確かに d が最大元となっている、つまり、

$$round_{\mathbb{DR}_f}(r) = d$$

となっている。

(c) $dnext_{\mathbb{DR}_f}(d) \leq r$ のとき

$$\{d' \in \mathbb{DR}_f \mid d' \leq r\}$$

には、$dnext_{\mathbb{DR}_f}(d)$ が含まれているので、d が最大元となることはない。

よって、$round_{\mathbb{DR}_f}(r) = d$ となるのは、$d \leq r < dnext_{\mathbb{DR}_f}(d)$ のときのみである。

(ii) $m^- \leq d \leq m^+$ のとき

$round_{\mathbb{DR}_f}$ の定義より、$r < m^-$ ならば、

$$round_{\mathbb{DR}_f}(r) < m^-$$

であり、$m^+ < r$ ならば、

$$round_{\mathbb{DR}_f}(r) > m^+$$

であるので、

$$m^- \leq round_{\mathbb{DR}_f}(r) \leq m^+$$

となりうるのは、$m^- \leq r \leq m^+$ のときに限られる。

(a) $0 \notin \mathbb{DR}_f$ のとき

d として考えられるのは、m^- と m^+ のみであるが、補題 5 の (ii)-(a) より、

$$round_{\mathbb{DR}_f}(r) = m^-$$

となるのは、

$$m^- \leq r < mid\left(m^-, m^+\right)$$

のときのみであり、

$$round_{\mathbb{DR}_f}(r) = m^+$$

となるのは、

$$mid\left(m^-, m^+\right) \leq r \leq m^+$$

のときのみである。

(b) $0 \in \mathbb{DR}_f$ のとき

d として考えられるのは、m^- と 0 と m^+ のみであるが、補題 5 の (ii)-(b) より、

$$round_{\mathbb{DR}_f}(r) = m^-$$

となるのは、

$$m^- \leq r < mid\left(m^-, 0\right)$$

のときのみであり、

$$round_{\mathbb{DR}_f}(r) = 0$$

となるのは、

$$mid\left(m^-, 0\right) \leq r \leq mid\left(0, m^+\right)$$

のときのみであり、

$$round_{\mathbb{DR}_f}(r) = m^+$$

となるのは、

$$mid\left(0, m^+\right) < r \leq m^+$$

のときのみである。

(iii) $m^+ < d$ のとき

$round_{\mathbb{DR}_f}$ の定義より、$m^+ < d$ ならば、

$$m^+ < round_{\mathbb{DR}_f}(d)$$

であり、$r \leq m^+$ ならば、

$$round_{\mathbb{DR}_f}(r) \leq m^+ < d$$

であるので、

$$round_{\mathbb{DR}_f}(r) = d$$

となるためには、$m^+ < r$ が必要条件となる。

(a) $r \leq dprev_{\mathbb{DR}_f}(d)$ のとき

$$\{d' \in \mathbb{DR}_f \mid r \leq d'\}$$

には、$dprev_{\mathbb{DR}_f}(d)$ が含まれているので、d が最小元となることはない。

(b) $dprev_{\mathbb{DR}_f}(d) < r \leq d$ のとき

このとき、

$$\left\{d' \in \mathbb{DR}_f \mid dprev_{\mathbb{DR}_f}(d) < r \leq d'\right\}$$

は d を含んでおり、かつ、$dprev_{\mathbb{DR}_f}$ の定義から、$\mathbb{DR}_f$ は $dprev_{\mathbb{DR}_f}(d)$ と d の間に元を持たないので、確かに d が最小元となっている、つまり、

$$round_{\mathbb{DR}_f}(r) = d$$

となっている。

(c) $d < r$ のとき

$$\{d' \in \mathbb{DR}_f \mid d < r \leq d'\}$$

の最小元は明らかに d より大きく、$round_{\mathbb{DR}_f}(r) = d$ となりえない。

よって、$round_{\mathbb{DR}_f}(r) = d$ となるのは、$dprev_{\mathbb{DR}_f}(d) < r \leq d$ のときのみである。

3.3.6 確率密度関数 f に従うことの証明

補題 5 で定義した $round_{\mathbb{DR}_f}$ が、任意の $d \in \mathbb{DR}_f$ に対して、(1) 式

$$P(d) = \int_{round^{-1}_{\mathbb{DR}_f}[\{d\}]} f(r)\,dr$$

を満たすことを証明すれば十分である。

証明の流れとしては、まず各 $d \in \mathbb{DR}_f$ に対して、そのような d を出力して停止するような、停止時の (r_{min}, r_{max}) として考えられうる全ての組を求める。その各組 (r_{min}, r_{max}) に対して、(r_{min}, r_{max}) で停止して d を出力する確率を、3.3.2.2 章にて述べた方法で求める。(これが $P(d)$ である。) 一方で、補題 6 で得られた $round^{-1}_{\mathbb{DR}_f}$ を用いて (1) 式の右辺を計算し、先ほど求めた $P(d)$ と一致することを確認する。

(i) $d < m^- < 0$ のとき

補題 3 より、$d < 0$ を出力して停止したとき、

$$r_{min} = d$$

となる。また、$dnext_{\mathbb{DR}_f}(d) \in \mathbb{DR}_f$ であることから、補題 2 の (3) より、

$$\begin{aligned} d &= r_{min} \\ &< r_{max} \\ &\leq dnext_{\mathbb{DR}_f}(d) \end{aligned}$$

となるが、ここで、$r_{max} < dnext_{\mathbb{DR}_f}(d)$ であるならば、補題 2 の (2) より、

$$\begin{aligned} d &= r_{min} \\ &< r_{max} \in \mathbb{DR}_f \\ &< dnext_{\mathbb{DR}_f}(d) \end{aligned}$$

となり、$dnext_{\mathbb{DR}_f}$ の定義に矛盾する。
ゆえに、RDG が d を出力したとき、停止時の (r_{min}, r_{max}) は $\left(d, dnext_{\mathbb{DR}_f}(d)\right)$ のみである。
よって、3.3.2.2 章より、

$$\begin{aligned} P(d) &= F(r_{max}) - F(r_{min}) \\ &= F\left(dnext_{\mathbb{DR}_f}(d)\right) - F(d) \\ &= \int_d^{dnext_{\mathbb{DR}_f}(d)} f(r)\,dr \end{aligned}$$

となり、補題 6 から、

$$\int_{round^{-1}_{\mathbb{DR}_f}[\{d\}]} f(r)\,dr = \int_d^{dnext_{\mathbb{DR}_f}(d)} f(r)\,dr$$

が得られるので、(1) 式が成立する。

(ii) $m^- \leq d \leq m^+$ のとき

(a) $0 \notin \mathbb{DR}_f$ のとき

d として考えられるのは、m^- と m^+ のみである。

(1) $d = m^- < 0$ のとき

補題 3 より、$m^- < 0$ を出力して停止したとき、

$$r_{min} = m^-$$

となる。また、$m^+ \in \mathbb{DR}_f$ であることから、補題 2 の (3) より、

$$\begin{aligned} m^- &= r_{min} \\ &< r_{max} \\ &\leq m^+ \end{aligned}$$

となるが、ここで、$r_{max} < m^+$ であるならば、補題 2 の (2) より、

$$\begin{aligned} m^- &= r_{min} \\ &< r_{max} (\neq 0) \in \mathbb{DR}_f \\ &< m^+ \end{aligned}$$

となり、m^- または m^+ の定義に矛盾する。
ゆえに、RDG が m^- を出力したとき、停止時の (r_{min}, r_{max}) は (m^-, m^+) のみである。
よって、3.3.2.2 章より、

$$\begin{aligned} P\left(m^-\right) &= \frac{F(r_{max}) - F(r_{min})}{2} \\ &= \frac{F(m^+) - F(m^-)}{2} \end{aligned}$$

となり、補題 6 から、

$$\begin{aligned} \int_{round_{\mathbb{DR}_f}} f(r)\,dr &= \int_{m^-}^{mid\left(m^-, m^+\right)} f(r)\,dr \\ &= F\left(mid\left(m^-, m^+\right)\right) - F\left(m^-\right) \\ &= F\left(F^{-1}\left(\frac{F(m^-) + F(m^+)}{2}\right)\right) \\ &\quad - F\left(m^-\right) \\ &= \frac{F(m^-) + F(m^+)}{2} - F\left(m^-\right) \\ &= \frac{F(m^+) - F(m^-)}{2} \end{aligned}$$

が得られるので、(1) 式が成立する。

(2) $d = m^+ > 0$ のとき

補題 3 より、$m^+ > 0$ を出力して停止したとき、

$$r_{max} = m^+$$

となる。また、$m^- \in \mathbb{DR}_f$ であることから、補題 2 の (3) より、

$$\begin{aligned} m^- &\leq r_{min} \\ &< r_{max} = m^+ \end{aligned}$$

となるが、ここで、$m^- < r_{min}$ であるならば、

補題 2 の (2) より、

$$\begin{aligned} m^- &< r_{min}(\neq 0) \in \mathbb{DR}_f \\ &< r_{max} = m^+ \end{aligned}$$

となり、m^- または m^+ の定義に矛盾する。
ゆえに、RDG が m^+ を出力したとき、停止時の (r_{min}, r_{max}) は (m^-, m^+) のみである。
よって、3.3.2.2 章より、

$$\begin{aligned} P\left(m^+\right) &= \frac{F\left(r_{max}\right) - F\left(r_{min}\right)}{2} \\ &= \frac{F\left(m^+\right) - F\left(m^-\right)}{2} \end{aligned}$$

となり、補題 6 から、

$$\begin{aligned} \int_{round_{\mathbb{DR}_f}} f(r)\,dr &= \int_{mid(m^-,m^+)}^{m^+} f(r)\,dr \\ &= F\left(m^+\right) - F\left(mid\left(m^-, m^+\right)\right) \\ &= F\left(m^+\right) - \frac{F\left(m^-\right) + F\left(m^+\right)}{2} \\ &= \frac{F\left(m^+\right) - F\left(m^-\right)}{2} \end{aligned}$$

が得られるので、(1) 式が成立する。

(b) $0 \in \mathbb{DR}_f$ のとき
d として考えられるのは、m^- と 0 と m^+ のみである。

(1) $d = m^- < 0$ のとき
補題 3 より、$m^- < 0$ を出力して停止したとき、

$$r_{min} = m^-$$

となる。また、$0 \in \mathbb{DR}_f$ であることから、補題 2 の (3) より、

$$\begin{aligned} m^- &= r_{min} \\ &< r_{max} \\ &\leq 0 \end{aligned}$$

となるが、ここで、$r_{max} < 0$ であるならば、補題 2 の (2) より、

$$\begin{aligned} m^- &= r_{min} \\ &< r_{max} \in \mathbb{DR}_f \\ &< 0 \end{aligned}$$

となり、m^- の定義に矛盾する。
ゆえに、RDG が m^- を出力したとき、停止時の (r_{min}, r_{max}) は $(0, m^-)$ のみである。
よって、3.3.2.2 章より、

$$\begin{aligned} P\left(m^-\right) &= \frac{F\left(r_{max}\right) - F\left(r_{min}\right)}{2} \\ &= \frac{F(0) - F\left(m^-\right)}{2} \end{aligned}$$

となり、補題 6 より、

$$\begin{aligned} \int_{round_{\mathbb{DR}_f}} f(r)\,dr &= \int_{m^-}^{mid\left(m^-,0\right)} f(r)\,dr \\ &= F\left(mid\left(m^-, 0\right)\right) - F\left(m^-\right) \\ &= F\left(F^{-1}\left(\frac{F\left(m^-\right) + F(0)}{2}\right)\right) \\ &\quad - F\left(m^-\right) \\ &= \frac{F\left(m^-\right) + F(0)}{2} - F\left(m^-\right) \\ &= \frac{F(0) - F\left(m^-\right)}{2} \end{aligned}$$

が得られるので、(1) 式が成立する。

(2) $d = 0$ のとき
補題 3 より、0 を出力して停止するのは、擬似コードの 31 にて RBG が 0 を出力した場合と 1 を出力した場合において、各々 1 通りずつの合計 2 通りが考えられる。
そこで、それぞれの場合について、停止時の (r_{min}, r_{max}) を求めていく。

- RBG が 0 を出力した結果として 0 が出力された場合

$$r_{min} = 0$$

となる。また、$m^+ \in \mathbb{DR}_f$ であることから、補題 2 の (3) より、

$$\begin{aligned} 0 = r_{min} \\ < r_{max} \\ \leq m^+ \end{aligned}$$

となるが、ここで、$r_{max} < m^+$ であるならば、補題 2 の (2) より、

$$\begin{aligned} 0 = r_{min} \\ < r_{max} \in \mathbb{DR}_f \\ < m^+ \end{aligned}$$

となり、m^+ の定義に矛盾する。
ゆえに、この場合においては、停止時の (r_{min}, r_{max}) は $(0, m^+)$ のみである。

- RBG が 1 を出力した結果として 0 が出力された場合

$$r_{max} = 0$$

となる。また、$m^- \in \mathbb{DR}_f$ であることから、補題 2 の (3) より、

$$\begin{aligned} m^- &\leq r_{min} \\ &< r_{max} = 0 \end{aligned}$$

となるが、ここで、$m^- < r_{min}$ であるならば、補題 2 の (2) より、

$$\begin{aligned} m^- &< r_{min} \in \mathbb{DR}_f \\ &< r_{max} = 0 \end{aligned}$$

となり、m^- の定義に矛盾する。
ゆえに、この場合においては、停止時の (r_{min}, r_{max}) は $(m^-, 0)$ のみである。

以上、RDG が 0 を出力したとき、停止時の (r_{min}, r_{max}) は $(m^-, 0)$ と $(0, m^+)$ のみである。
よって、3.3.2.2 章より、

$$\begin{aligned} P(0) &= \frac{F(0) - F(m^-)}{2} + \frac{F(m^+) - F(0)}{2} \\ &= \frac{F(m^+) - F(m^-)}{2} \end{aligned}$$

となり、補題 6 より、

$$\begin{aligned} \int_{round_{\mathbb{DR}_f}} f(r)\,dr &= \int_{mid(m^-,0)}^{mid(0,m^+)} f(r)\,dr \\ &= F\left(mid\left(0, m^+\right)\right) \\ &\quad - F\left(mid\left(m^-, 0\right)\right) \\ &= F\left(F^{-1}\left(\frac{F(0) + F(m^+)}{2}\right)\right) \\ &\quad - \frac{F(m^-) + F(0)}{2} \\ &= \frac{F(0) + F(m^+)}{2} \\ &\quad - \frac{F(m^-) + F(0)}{2} \\ &= \frac{F(m^+) - F(m^-)}{2} \end{aligned}$$

が得られるので、(1) 式が成立する。

(3) $d = m^+ > 0$ のとき
補題 3 より、$m^+ > 0$ を出力して停止したとき、

$$r_{max} = m^+$$

となる。また、$0 \in \mathbb{DR}_f$ であることから、補題 2 の (3) より、

$$\begin{aligned} 0 &\le r_{min} \\ &< r_{max} = m^+ \end{aligned}$$

となるが、ここで、$0 < r_{min}$ であるならば、補題 2 の (2) より、

$$\begin{aligned} 0 &< r_{min} \in \mathbb{DR}_f \\ &< r_{max} = m^+ \end{aligned}$$

となり、m^+ の定義に矛盾する。
ゆえに、RDG が m^+ を出力したとき、停止時の (r_{min}, r_{max}) は $(0, m^+)$ のみである。
よって、3.3.2.2 章より、

$$\begin{aligned} P\left(m^+\right) &= \frac{F(r_{max}) - F(r_{min})}{2} \\ &= \frac{F(m^+) - F(0)}{2} \end{aligned}$$

となり、補題 6 より、

$$\begin{aligned} \int_{round_{\mathbb{DR}_f}} f(r)\,dr &= \int_{mid(0,m^+)}^{m^+} f(r)\,dr \\ &= F\left(m^+\right) - F\left(mid\left(0, m^+\right)\right) \\ &= F\left(m^+\right) - \frac{F(0) + F(m^+)}{2} \\ &= \frac{F(m^+) - F(0)}{2} \end{aligned}$$

が得られるので、(1) 式が成立する。

(iii) $0 < m^+ < d$ のとき
補題 3 より、$d > 0$ を出力して停止したとき、

$$r_{max} = d$$

となる。また、$dprev_{\mathbb{DR}_f}(d) \in \mathbb{DR}_f$ であることから、補題 2 の (3) より、

$$\begin{aligned} dprev_{\mathbb{DR}_f}(d) &\le r_{min} \\ &< r_{max} = d \end{aligned}$$

となるが、ここで、$dprev_{\mathbb{DR}_f}(d) < r_{min}$ であるならば、補題 2 の (2) より、

$$\begin{aligned} dprev_{\mathbb{DR}_f}(d) &< r_{min} \in \mathbb{DR}_f \\ &< r_{max} = d \end{aligned}$$

となり、$dprev_{\mathbb{DR}_f}$ の定義に矛盾する。
ゆえに、RDG が d を出力したとき、停止時の (r_{min}, r_{max}) は $\left(dprev_{\mathbb{DR}_f}(d), d\right)$ のみである。
よって、3.3.2.2 章より、

$$\begin{aligned} P(d) &= F(r_{max}) - F(r_{min}) \\ &= F(d) - F\left(dprev_{\mathbb{DR}_f}(d)\right) \\ &= \int_{dprev_{\mathbb{DR}_f}(d)}^{d} f(r)\,dr \end{aligned}$$

となり、補題 6 から、

$$\int_{round^{-1}_{\mathbb{DR}_f}[\{d\}]} f(r)\,dr = \int_{dprev_{\mathbb{DR}_f}(d)}^{d} f(r)\,dr$$

が得られるので、(1) 式が成立する。

4. 実験

4.1 目的

ここまでで、提案手法の正当性を実数上において証明した。この章では、単純に IEEE754 倍精度浮動小数点数を用いた場合の挙動について調べる。

4.2 方法

4.2.1 乱数生成器

- 提案手法
  (b,h) 間隔条件を満たすよう指定した RDG を、$RDG\,(b,h)$ と呼ぶこととする。なお、今回は IEEE754 倍精度浮動小数点数を用いることに対応して、$b = B = 2^{-1022}$ とした。
- 逆変換法
  比較相手として、当アルゴリズムの基となっている、逆変換法を用いる。逆変換法とは、$[0,1]$ 上の一様乱数 u を用いて、$F\,(r) = u$ を満たす r を求めることにより、希望の分布に従う乱数 r を得る方法である。ここで、仮定より F は単調増加関数であるので、r は一意に存在する。
  なお、n ビットの一様乱数 $U\,(n)$、すなわち、
  $$U\,(n) = \left\{ \frac{k}{2^n+1} \mid 1 \leq k \in \mathbb{N} \leq 2^n \right\}$$
  を一様乱数として用いる逆変換法乱数生成器を、$ITM\,(n)$ と呼ぶこととする。(ITM = Inverse Transformation Method)

4.2.2 分布

今回使用した確率分布を表 1 に示す。

分布名	確率密度関数	累積分布関数逆関数
ラプラス分布 $(r<0)$	$\frac{e^r}{2}$	$r = \log(2u)$
ラプラス分布 $(0 \leq r)$	$\frac{e^{-r}}{2}$	$r = -\log(2(1-u))$
ロジスティック分布	$\frac{e^{-r}}{(1+e^{-r})^2}$	$r = -\log\left(\frac{1}{u}-1\right)$
コーシー分布	$\frac{1}{\pi(r^2+1)}$	$r = \tan\left(\pi\left(u-\frac{1}{2}\right)\right)$

表 1 確率分布

4.2.3 データ

乱数生成器 $RNG(RNG \in \{RDG\,(B,h)\,, ITM\,(n)\})$ による生成される乱数の集合を $RAND_{RNG}$ とおき、$RAND_{RNG}$ の最小元と最大元をそれぞれ $RAND_{MIN}$ と $RAND_{MAX}$ とおく。このとき、各 $d \in RAND_{RNG}$ に対して、d の両隣の元との相対間隔 $Dist\,(d)$ を、次のように定義する。

$$Dist\,(d) = \begin{cases} \frac{\frac{next(d)-d}{next(d)} + \frac{d-prev(d)}{prev(d)}}{2} = 1 & d = 0 \\ \frac{next(d)-d}{d} & d = RAND_{MIN} \\ \frac{d-prev(d)}{d} & d = RAND_{MAX} \\ \frac{\frac{next(d)-d}{d} + \frac{d-prev(d)}{d}}{2} = \frac{next(d)-prev(d)}{2d} & \text{それ以外} \end{cases}$$

ただし、

$$prev\,(d) : \{r \in RAND_{RNG} \mid r < d\} \text{の最大元}$$
$$next\,(d) : \{r \in RAND_{RNG} \mid d < r\} \text{の最小元}$$

である。

以上の準備の上で、次の二つをデータとして示す。まず乱数生成範囲として、$RAND_{MIN}$ と $RAND_{MAX}$ を表 (表 3、表 5、表 7) に示す。次に出現値間隔として、横軸に $d \in RAND_{RNG}$ を、縦軸に $Dist\,(d)$ をプロットしたものを、図 (図 1、図 3、図 5) に示す。

4.3 結果と議論

図 5、表 7 から、コーシー分布だけ他の分布と異なった挙動を取っていることが分かる。これは、コーシー分布と他の分布では、累積分布関数の逆関数が 0 と 1 付近で取る挙動が異なっていることによるものと思われる。例えば、

$$\lim_{u \to +0} \frac{F^{-1}\,(u)}{\frac{1}{u}}$$

が、コーシー分布だけ負の値 $(\frac{-1}{\pi})$ に収束するが、他の分布では 0 に収束する。さらに、実数上では $F^{-1}\,(0)$ と $F^{-1}\,(1)$ は本来無限大になるはずだが、浮動小数点数で計算すると、表 2 の通り、コーシー分布だけ有限の値となることも原因の一つであると考えられる。

また、図 1、図 3、表 3、表 5 から、ラプラス分布とロジスティック分布では、左側のテールのみ改善されており、右側のテールはほとんど改善されていないことが分かる。これは、浮動小数点数が、0 付近と 1 付近で表現可能な値の間隔が異なることが原因であると考えられる。例えば、0 の次の浮動小数点数は $2^{-1022-52}$ であるが、1 の前の浮動小数点数は $1-2^{-52}$ であり、絶対間隔の差は 2^{1022} 倍である。

これらのことから、提案手法は、単純に浮動小数点数へ適用することができず、実装のためには、更なる議論が必要であることが分かる。

分布名	$F^{-1}\,(0)$	$F^{-1}\,(1)$
ラプラス分布	$-\infty$	∞
ロジスティック分布	$-\infty$	∞
コーシー分布	-1.633124×10^{16}	1.633124×10^{16}

表 2 計算機上での $F^{-1}\,(0)$ と $F^{-1}\,(1)$ の計算結果

4.4 対策

右側のテールが改善されないことへの対策として、内部の演算において倍倍精度 [16–18] を用いた場合の結果を、表 (表 4、表 6、表 8) と図 (図 2、図 4、図 6) に示す。

全ての分布において、右側のテールが改善していることが分かる。特に、図 6 から、コーシー分布においては、既存逆変換法では、倍倍精度を用いても左側のテールしか改善しないが、提案手法においては右側のテールも改善している。

これらのことは、多倍長や可変長の浮動小数を用いることで、提案手法が実数上のみならず利用可能である可能性を示唆しているものと思われる。

乱数生成器	$RAND_{MIN}$	$RABD_{MAX}$
$RDG\,(B, B \times 10^{-4})$	-743.1	35.35
$RDG\,(B, B \times 10^{-8})$	-743.1	35.35
$RDG\,(B, B \times 10^{-16})$	-743.1	35.35
$ITM\,(32)$	-21.49	21.49
$ITM\,(64)$	-43.67	36.04
$ITM\,(1075)$	-743.7	36.04

表 3 乱数生成範囲 (ラプラス分布)：倍精度使用

乱数生成器	$RAND_{MIN}$	$RABD_{MAX}$
$RDG\,(B, B \times 10^{-4})$	-743.1	743.1
$RDG\,(B, B \times 10^{-8})$	-743.1	743.1
$RDG\,(B, B \times 10^{-16})$	-743.1	743.1
$ITM\,(32)$	-21.49	21.49
$ITM\,(64)$	-43.67	36.04
$ITM\,(1075)$	-743.7	36.04

表 4 乱数生成範囲 (ラプラス分布)：倍倍精度使用

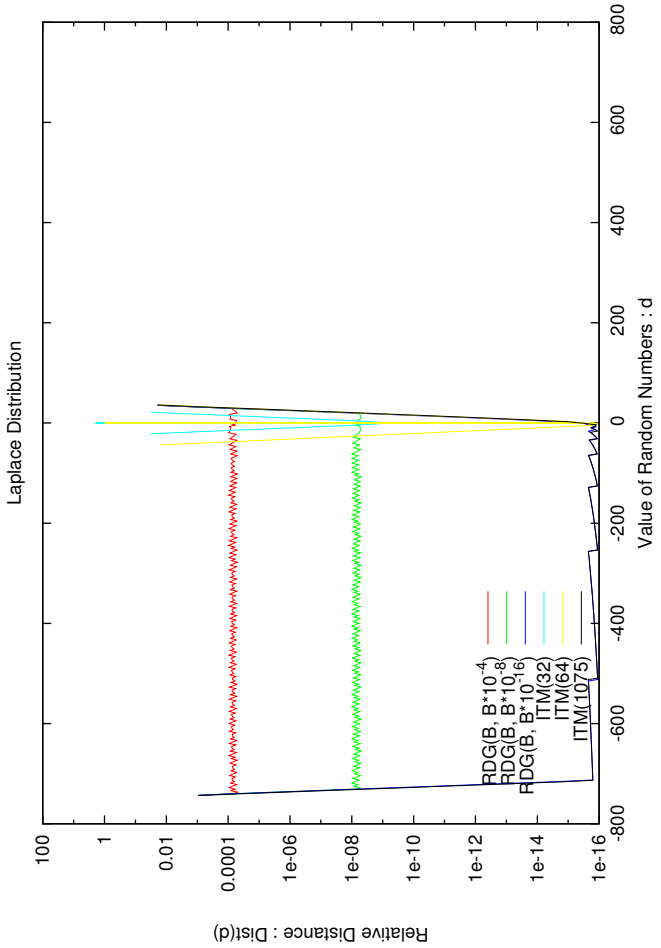

図 1 相対間隔 (ラプラス分布)：倍精度使用

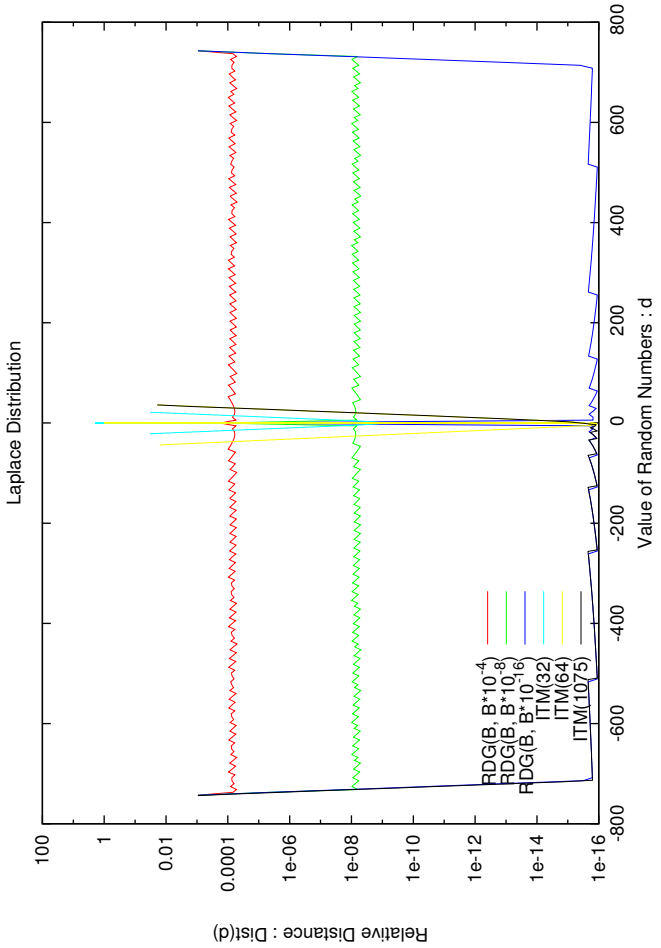

図 2 相対間隔 (ラプラス分布)：倍倍精度使用

乱数生成器	$RAND_{MIN}$	$RABD_{MAX}$
$RDG\,(B, B \times 10^{-4})$	-709.8	36.04
$RDG\,(B, B \times 10^{-8})$	-709.8	36.04
$RDG\,(B, B \times 10^{-16})$	-709.8	36.04
$ITM\,(32)$	-22.18	22.18
$ITM\,(64)$	-44.36	36.04
$ITM\,(1075)$	-709.8	36.04

表 5 乱数生成範囲 (ロジスティック分布)：倍精度使用

乱数生成器	$RAND_{MIN}$	$RABD_{MAX}$
$RDG\,(B, B \times 10^{-4})$	-691.1	743.7
$RDG\,(B, B \times 10^{-8})$	-691.1	743.7
$RDG\,(B, B \times 10^{-16})$	-691.1	743.7
$ITM\,(32)$	-22.18	22.18
$ITM\,(64)$	-44.36	36.74
$ITM\,(1075)$	-691.1	36.74

表 6 乱数生成範囲 (ロジスティック分布)：倍倍精度使用

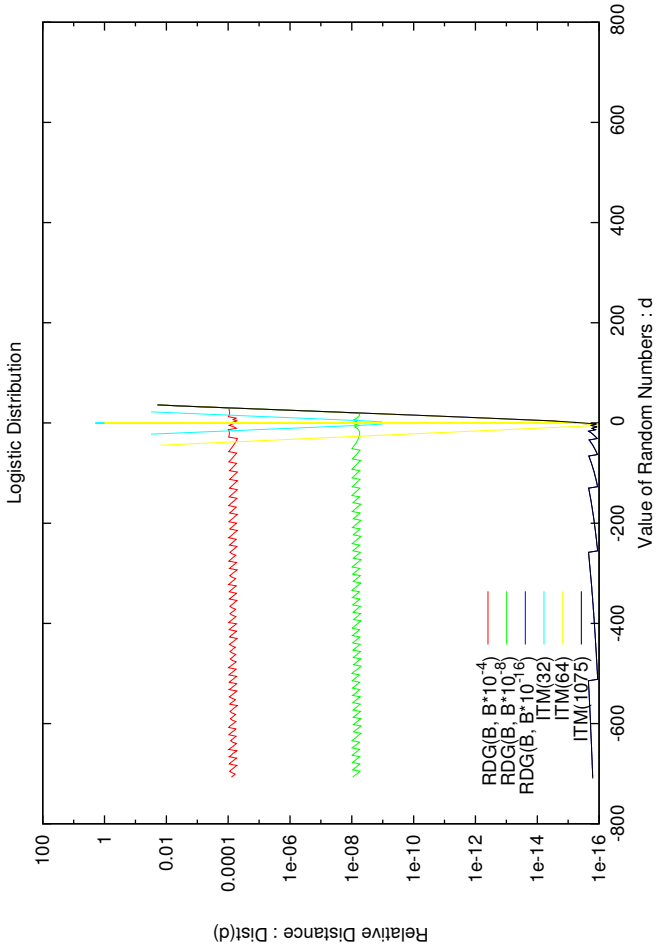

図 3 相対間隔 (ロジスティック分布)：倍精度使用

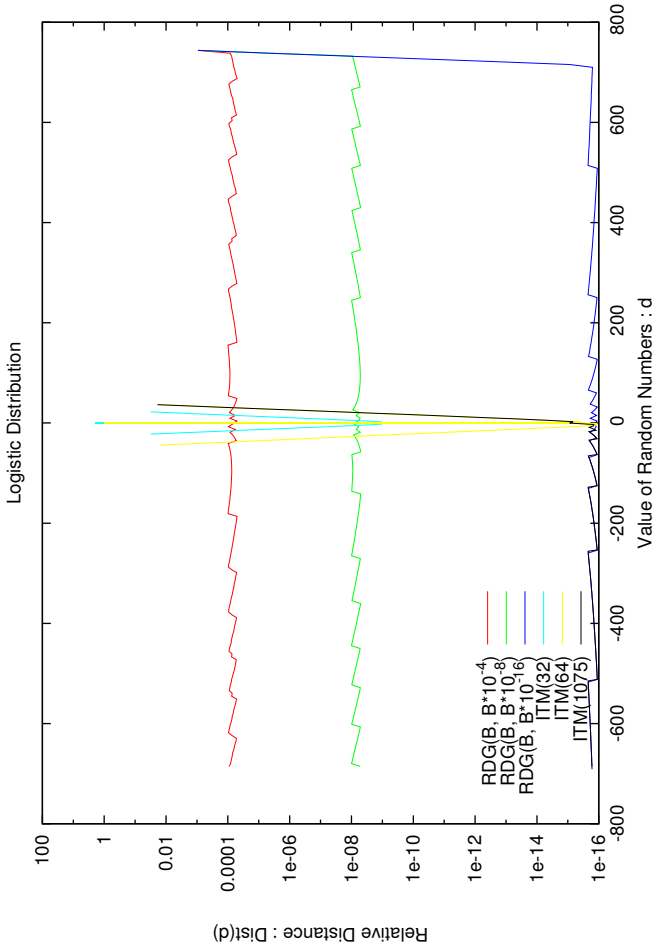

図 4 相対間隔 (ロジスティック分布)：倍倍精度使用

乱数生成器	$RAND_{MIN}$	$RABD_{MAX}$
$RDG\,(B, B \times 10^{-4})$	-1.633×10^{16}	1.375×10^{15}
$RDG\,(B, B \times 10^{-8})$	-1.633×10^{16}	1.375×10^{15}
$RDG\,(B, B \times 10^{-16})$	-1.633×10^{16}	1.375×10^{15}
$ITM\,(32)$	-1.367×10^{9}	1.367×10^{9}
$ITM\,(64)$	-1.633×10^{16}	1.633×10^{16}
$ITM\,(1075)$	-1.633×10^{16}	1.633×10^{16}

表 7 乱数生成範囲 (コーシー分布)：倍精度使用

乱数生成器	$RAND_{MIN}$	$RABD_{MAX}$
$RDG\,(B, B \times 10^{-4})$	-1.633×10^{16}	1.633×10^{16}
$RDG\,(B, B \times 10^{-8})$	-1.633×10^{16}	1.633×10^{16}
$RDG\,(B, B \times 10^{-16})$	-1.633×10^{16}	1.633×10^{16}
$ITM\,(32)$	-1.367×10^{9}	1.367×10^{9}
$ITM\,(64)$	-1.629×10^{16}	1.633×10^{16}
$ITM\,(1075)$	-1.633×10^{16}	1.633×10^{16}

表 8 乱数生成範囲 (コーシー分布)：倍倍精度使用

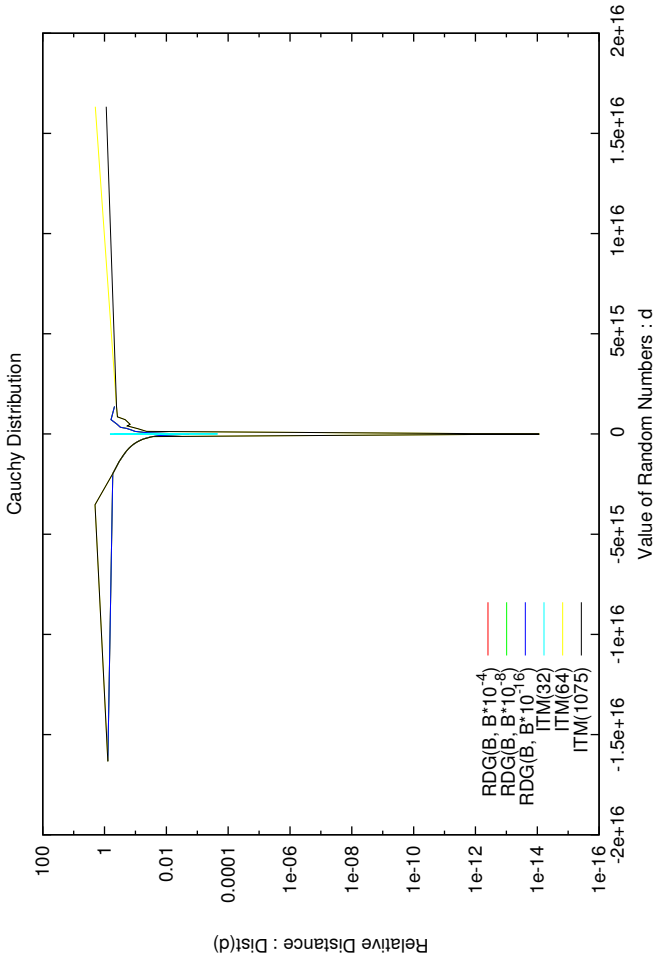

図 5 相対間隔 (コーシー分布)：倍精度使用

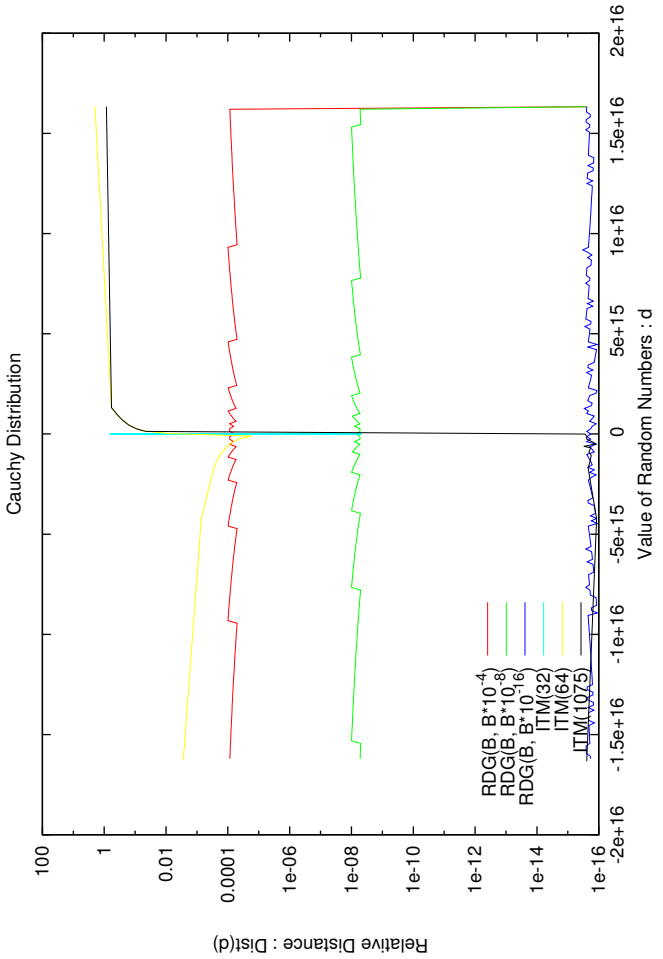

図 6 相対間隔 (コーシー分布)：倍倍精度使用

5. 関連研究

- テール領域の精度向上に関する取り組み
  テール領域の精度向上に関しては、1.1 章でも取り上げたが、正規分布のテール領域における精度向上への取り組みが挙げられる [14]。この論文では、主には正規乱数生成器の紹介や分類、性能評価を行っているが、その一部として、これまであまり注目されることのなかったテール領域の精度に着目し、一様整数乱数を浮動小数点数へ変換する際の方法について述べている。
- テール領域の乱数生成に関する取り組み
  テール領域の乱数生成に関しては、1.1 章でも取り上げたが、テール領域を他の分布で近似すること無く、高速な乱数生成を目指す取り組みが挙げられる [15]。この論文では、単位正方形を確率密度関数の分位へ投影する方法に対して、単位正方形を部分正方形へ分割し、それぞれの部分正方形に事前計算を行うことで、高速な乱数生成を達成している。
- テール領域の危険性に関する考察
  テール領域の危険性に関して、パレート分布を例に考察したものが挙げられる [19]。テール領域は絶対値の大きい値で構成されていることから、出現確率が低くともシミュレーション結果に強い影響力を持っているが、中にはシミュレーション対象にとって現実離れした値も含まれている。そのため、この論文では、シミュレーションの長さに応じてテール領域を制限することを提案している。
- 分布の離散化に関する取り組み (モーメント保持)
  分布の離散化に関して、与えられたモーメントを保存するような方法が挙げられる [20]。一般に、離散化した点の数はモーメントの制約による等式よりもはるかに多く、一意に解けないことが多い。そのため、この論文では、最大エントロピー原理 [21] に基づいて、シャノンの情報量 [22] を最大化することで、分布を近似している。その最大化における制約こそが、与えられたモーメントとなっている。
- 分布の離散化に関する取り組み (累積分布関数保持)
  分布の離散化に関するもう一つの方法として、モーメントだけではなく、累積分布関数も (いくつかの点において) 一致させるような方法がある [23]。この論文では、$(2N-1)$ 次以下の全てのモーメントと、少なくとも $(M+1)$ 点における累積分布関数が、元の分布と一致するような、$(M \times N)$ 点による離散化について述べられている。

6. 結論

この論文では、テールの精度保証を目的として、

- テールの精度の定義
- テールの精度向上手法の提案
- 精度保証の実数上における証明

を行った。当研究の制約としては、

(1) 単純には浮動小数点数へ適用できない
(2) 確率密度関数の正値性を仮定した
(3) 累積分布関数の逆関数を必要とする

等が挙げられる。よって、今後の課題として、

(1) 浮動小数点数における実装を議論する
(2) 定義や証明を修正して任意の確率密度関数に対する議論をする
(3) 累積分布関数の逆関数を必要としない方法を考える

等が挙げられる。

参考文献

[1] Salmon, J. K., Moraes, M. A., Dror, R. O. and Shaw, D. E.: Parallel Random Numbers: As Easy As 1, 2, 3, *Proceedings of 2011 International Conference for High Performance Computing, Networking, Storage and Analysis*, SC '11, New York, NY, USA, ACM, pp. 16:1–16:12 (2011).
[2] Matsumoto, M. and Nishimura, T.: Mersenne Twister: A 623-dimensionally Equidistributed Uniform Pseudo-random Number Generator, *ACM Trans. Model. Comput. Simul.*, Vol. 8, No. 1, pp. 3–30 (1998).
[3] Payne, W. H., Rabung, J. R. and Bogyo, T. P.: Coding the Lehmer Pseudo-random Number Generator, *Commun. ACM*, Vol. 12, No. 2, pp. 85–86 (1969).
[4] Box, G. E. P. and Muller, M. E.: A Note on the Generation of Random Normal Deviates, *The Annals of Mathematical Statistics*, Vol. 29, No. 2, pp. 610–611 (1958).
[5] Bell, J. R.: Algorithm 334: Normal Random Deviates, *Commun. ACM*, Vol. 11, No. 7, pp. 498– (1968).
[6] Knop, R.: Remark on Algorithm 334 [G5]: Normal Random Deviates, *Commun. ACM*, Vol. 12, No. 5, pp. 281– (1969).
[7] Kabal, P.: Generating Gaussian Pseudo-Random Deviates, Technical report, Department of Electrical and Computer Engineering McGill University (2000).
[8] Marsaglia, G. and Tsang, W. W.: The Ziggurat Method for Generating Random Variables, *Journal of Statistical Software*, Vol. 5, No. 8, pp. 1–7 (2000).
[9] Marsaglia, G. and Tsang, W. W.: The Monty Python Method for Generating Random Variables, *ACM Trans. Math. Softw.*, Vol. 24, No. 3, pp. 341–350 (1998).
[10] Marsaglia, G. and Bray, T. A.: A Convenient Method for Generating Normal Variables, *SIAM Review*, Vol. 6, No. 3, pp. pp. 260–264 (1964).
[11] Brent, R. P.: Algorithm 488: A Gaussian Pseudo-random Number Generator, *Commun. ACM*, Vol. 17, No. 12, pp. 704–706 (1974).
[12] Kinderman, A. J. and Monahan, J. F.: Computer Generation of Random Variables Using the Ratio of Uniform Deviates, *ACM Trans. Math. Softw.*, Vol. 3, No. 3, pp. 257–260 (1977).
[13] Panjer, H. H.: *Operational Risk: Modeling Analytics*, pp. 411–414, Wiley-Interscience (2006).
[14] Thomas, D. B., Luk, W., Leong, P. H. and Villasenor, J. D.: Gaussian Random Number Generators, *ACM Comput. Surv.*, Vol. 39, No. 4 (2007).
[15] Fulger, D., Scalas, E. and Germano, G.: Random num-

bers from the tails of probability distributions using the transformation method, *CoRR*, Vol. abs/0902.3207 (2009).

[16] Dekker, T.: A floating-point technique for extending the available precision, *Numerische Mathematik*, Vol. 18, No. 3, pp. 224–242 (1971).

[17] Hida, Y., Li, X. S. and Bailey, D. H.: Algorithms for Quad-Double Precision Floating Point Arithmetic, *Proceedings of the 15th IEEE Symposium on Computer Arithmetic*, ARITH '01, Washington, DC, USA, IEEE Computer Society, pp. 155– (2001).

[18] 小武守恒，藤井昭宏，長谷川秀彦， 西田晃：倍精度と 4 倍精度の混合型反復法の提案，2007 年ハイパフォーマンスコンピューティングと計算科学シンポジウム論文集，pp. 9–16 (2007).

[19] Feitelson, D. G.: Random Number Generators and Heavy-Tail Distributions.

[20] TanakaKen' ichiro，Toda, A. A. : Discrete approximations of continuous distributions by maximum entropy, *Economics Letters*, Vol. 118, No. 3, pp. 445–450 (2013).

[21] Jaynes, E. T.: Information Theory and Statistical Mechanics, *Physical Review*, Vol. 106(4), pp. 620–630 (1957).

[22] Shannon, C. E.: A mathematical theory of communication, *Bell Systems Technical Journal*, Vol. 27, pp. 379–423,623–656 (1948).

[23] Luceño, A.: Discrete approximations to continuous univariate distributions-an alternative to simulation, *Journal of the Royal Statistical Society Series B*, Vol. 61, No. 2, pp. 345–352 (1999).