情報処理
第2回 コンピュータにおける数の世界
2000年4月19日

1　コンピュータにおける数の世界

コンピュータが取り扱える数値は,

(1) 整数型 (10 進数表示): 0, 100, -10, など 符号 + 数値の大きさ.
(2) 実数型　(10 進数表示): 0.001, 3.14, -1.234×10^2 など
符号あるいは小数点 + 数値の大きさ.

1.1　整数型数値

コンピュータの中では 2 進数に変換されて記憶される. 例えば, $4_{10} \longrightarrow 100_2$. 2 進数の各行をビット (bit) といい, 2 の補数を用いてマイナスの符号を決めている.

1. (i) 短い整数型 (short): 符号をあわせて 16 ビット.

$$\begin{array}{rccc} \text{範囲}: & 1000000000000000_{\,2} & \sim & 0111111111111111_{\,2} \\ & \| & & \| \\ & (-2^{15})_{10} = -32768 & \sim & (2^{15}-1)_{10} = 32767 \end{array}$$

(ii) 長い整数型 (long): 符号をあわせて 32 ビット.

$$\text{範囲}: \quad (-2^{31})_{10} = -2147483648 \quad \sim \quad (2^{31}-1)_{10} = 2147483647$$

注意 32 ビットのコンピュータのとき,

$$1111111111111111_{\,2} + 0000000000000001_{\,2} = 0000000000000000_{\,2}$$

となるので (17 ビット以降は無視), $1111111111111111_{\,2} = (-1)_{10}$. よって,
$1000000000000000_{\,2} = (-1-(2^{15}-1))_{10} = (-2^{15})_{10}$ となる.

2　2の補数

二進法で負の数を表示するときは次のようになる.
4 ビットの場合:

2 進法表示	10 進法表示
0001	1
0010	2
0011	3
0100	4
0101	5
0110	6
0111	7
0000	0
1111	-1
1110	-2
1101	-3
1100	-4
1011	-5
1010	-6
1001	-7
1000	-8

負の求め方として補数の概念がある.

定義 1 *2* 進数の *1* の補数*:* 各桁を 1 としたものから与えられて数を引く.

例　4 ビットのコンピュータの場合 $(0001)_2$ の 1 の補数は

$$(1111)_2 - (0001)_2 = (1110)_2$$

となります.

定義 2 *2* 進数の *2* の補数*: 1* の補数に *1* を加えたもの.

例　4 ビットのコンピュータの場合
$(0001)_2$ の 2 の補数は

$$(1110)_2 + (0001)_2 = (1111)_2$$

となります.

$(0011)_2$ の 2 の補数は

(1) $(0011)_2$ の 1 の補数は

$$(1111)_2 - (0011)_2 = (1100)_2.$$

(2) よって 2 の補数は

$$(1100)_2 + (0001)_2 = (1101)_2$$

となり, これが 4 ビットの場合-3 を表している. (上の表をみよ)

3 n 進法

コンピュータで用いられている進法は 2 進法以外に, 8 進法, 16 進法があります. 2 進法と同様に a が n 進法で表示されているとき $(a)_n$ と書くとすると,

$$(a_n a_{n-1} \cdots a_1 a_0)_m = (a_n \times m^n + a_{n-1} \times m^{n-1} + \cdots + a_1 \times m + a_0)_{10}$$

となる.

- **2 進法と 8 進法**

  2 進法を 8 進法に変換する方法は, 3 ブロックごとに展開していく: $(11010110)_2$ の場合

  $$(11|\underbrace{010}_{3}|\underbrace{110}_{3})_2 \Longrightarrow (326)_8$$

  8 進法から 2 進法はこの逆を行う.

- **2 進法と 16 進法**

  2 進法から 16 進法に変換するとき 4 ブロックごとに展開していく.

- **16 進法と 8 進法**

  一旦 2 進法に変換してからそれぞれに変換していけばよい.

- 小数表示

  整数同様 n 進数の小数表示を

  $$(0.a_1 a_2 \cdots a_l)_n$$

  と表す.

  例

  $$(0.1011)_2 = (\tfrac{1}{2} + \tfrac{1}{2^3} + frac12^4)_{10} = (0.6875)_{10}$$
  $$(0.405)_8 = (\tfrac{1}{8} \times 4 + \tfrac{1}{8^3} \times 5)_{10} = (0.509765625)_{10}$$
  $$(0.3A)_{16} = (\tfrac{3}{16} + \tfrac{10}{16^2})_{10} = (0.2265125)_{10}$$

  例えば, 10 進法小数表示を 2 進法小数表示に変換する場合は与えられた数を 2 倍していき, 整数部分を順にとってくればよい.

例

$(0.1)_{10} = (0.a_1a_2\cdots a_n)_2$ 2 変換する場合は,

$$\begin{array}{lll} 0.1 \times 2^4 = 1.6 & \longrightarrow & a_1 = 0, a_2 = 0, a_3 = 0, a_4 = 1 \\ 0.6 \times 2 = 1.2 & \longrightarrow & a_5 = 1 \\ 0.2 \times 2 = 0.4 & \longrightarrow & a_6 = 0 \\ \cdots & & \cdots \end{array}$$

より

$$(0.1)_{10} = (0.000110001100011\cdots)_2$$

となる.

3.1 実数型整数値

計算機内で適当に切り上げて $\beta = 16$ 進法で

$$\pm(0.d_1d_2d_3d_4d_5d_6)_\beta \times \beta^b$$

で表示する. 但し,

$$\pm(0.d_1d_2d_3d_4d_5d_6)_\beta = \pm\left(\frac{d_1}{\beta} + \frac{d_2}{\beta^2} + \cdots + \frac{d_6}{\beta^6}\right)$$

$(0 < d_1 \leq 15,\ 0 \leq d_i \leq 15\ (\ i \geq 2),\ b$:7 ビット$)$. よって, 最大値は

$$\left(\frac{15}{16} + \frac{15}{16^2} + \cdots + \frac{15}{16^6}\right) \times 16^{63} :=: 17 \times 10^{75}$$

であり, これより値が大きいときはオーバーフローをおこし正しい計算ができない. 最小値は,

$$\frac{1}{16} \times 16^{-64} = 16^{-65} :=: 5 \times 10^{-79}$$

となり, これより小さい値に対してはアンダーフローをおこし 0 になってしまう.

このため計算の順序により値が異なってしまうときがある.

例

$$\begin{array}{lll} (10^{-50} \times 10^{-40}) \times 10^{60} & = 0 \times 10^{60} & = 0 \\ 10^{-50} \times (10^{-40} \times 10^{60}) & = 10^{-50} \times 10^{20} & = 10^{-30} \end{array}$$

上の表示は単精度の表示であり, 現在はその二倍の倍精度の表示が用いられている. 但し, 指数部分 (b) は同じで, 仮数部分 (d_i) は $14(56$ ビット$)$ ある.

4　誤差

数値計算では有限桁の数値だけを取り扱うので誤差 (真の数との差) が生ずる.

(1) 絶対誤差:
a を真の数, x をその近似値としたとき $e = x - a$ を誤差といい, $|e| = |x - a|$ を絶対誤差という.

(2) 相対誤差:
上の表現を用いて, 相対誤差 e_R を $e_R = \frac{e}{a} = \frac{x-a}{a}$ であらわす. 実用的には, $e_R = \frac{e}{x}$ を用いる.

(3) 許容誤差:
許容絶対誤差を ϵ_A, 許容相対誤差を ϵ_R としたとき, 近似値 x が下記の条件を満たしたとき真の値に十分近いと判断する:

$$\begin{aligned} |x - a| &< \epsilon_A \\ \frac{|x-a|}{|x|} &< \epsilon_R \end{aligned}$$

例えば, 2 つの地点 P, Q 間の距離の測定値が $100\mathrm{m}$, 別の 2 地点 R, S 間の距離の測定値が $42195\mathrm{m}$ であり, これらの誤差が共に $10\mathrm{cm}$ であったとします. このとき誤差の限界が同じでも近似の程度はかなり異なることは感覚的にわかると思います. 実際, 相対誤差評価をすると, それぞれ 0.1, 2.3×10^{-4} となり後ろのケースの法が精度は高いことがわかります.
ϵ_A として使用しているコンピュータで使用できる 0 以外の最小の絶対値数を採用し, ϵ_R として計算機イプシロンを採用してかまわない. 但し, 計算機イプシロンとは $1 + e \geq 1$ となる最小の e のことである. (プログラムを書いて RAINBOW のマシンの計算機イプシロンを自分で調べて見よ)

(3) 丸め誤差:
真の数を適当なところで四捨五入し切り捨て, 切り上げて生じた誤差. (β 進法のとき $\frac{\beta}{2}$ で入り, $\frac{\beta}{2} - 1$ で切り)

(i) 情報落ち:
絶対値の差が大きな 2 つの数の加減で小さな値が無視される.

(ii) 桁落ち:
減算や異符号で絶対値の近い数の加算をおこなった際に生じる有効桁数の減.

(iii) 積み残し:
情報落ちが原因で発生する現象.

(iv) 打ち切り誤差:
無限級数など有限個の和で打ち切った際に生じる誤差.

(V) 代入誤差:
関数 $f(x)$ に真の数 a の代わりに近似値 x を代入した際などに生じる誤差. (この場合は $f(x) - f(a)$)

演習問題 1

学籍番号
氏名

(1) $(01110101)_2$ に対する 2 の補数を求めよ.
(2) $(1.0625)_{10}$ を 2 進数表示に変換せよ.
(3) $(100)_{16}$ を 8 進数に変換せよ.
(4) 次の 16 進数の加算をせよ (結果も 16 進数表示).

$$(469)_{16} + (2AE)_{16}$$

解答欄