2.4 情報とデータ

「情報 (information)」という言葉は、現代社会において重要なキーワードのひとつとなっています。この言葉が最初に登場したのは、明治 9 年 (1876) で、フランス語の軍事用語 renseignement の訳語として、「敵情の報知 (報告)」を縮めて「情報[6]」と名付けられたとされています。その後、ラテン語で「明確な形を与えるもの」の意味の informare を語源とする information の訳語として「情報」が使われるようになりました。現在では、広辞苑などで「物事の事情, 状態についての知らせ」と説明され、必要とする人間にとって混沌としたデータの羅列ではなく明確な意味を持つものを「情報」と呼びます。すなわち、「特定の人に有用なデータ, 意味あるデータ」又は「データから取り出した意味あるもの」を示しています。

さて、これまで、「データ (data)」という言葉を断りなく使ってきましたが、上記を加味すると、この言葉は「観察等によって得た混沌とした事実」又は「処理されていない (意味づけされていない) 単なる事実」を示しています。すなわち、「情報」と「データ」には、

$$\text{「情報」} = \text{「データ」} + \text{意味}$$

という関係があり、「データ」に意味づけをしたものを「情報」と呼びます。1 章で述べたように、コンピュータは、データ (または情報) を処理してすることによって必要な結果 (意味づけされたデータ, 情報) を外部に返します。

今日では、文字・画像・音楽など様々なデータ (情報) が数値化され、コンピュータで扱うことができます。この情報を定量的に取り扱うことを初めて明確な形で提唱したのは、シャノン (C. E. Shannon, 1916–2001) で、1948 年に「Mathematical Theory of Communication」という情報理論に関する論文を発表しました。この論文の中で、情報量を表す単位として 2 進数 1 桁という意味を持つビット (bit: binary digit) が、初めて導入されました。

情報量　事象 A が起こる確率を $P(A)$ としたとき情報量 $I(A)$ は

$$I(A) = -\log_2 P(A)$$

によって定義されます。特に、情報科学では対数の底を 2 にとり、1 ビットの情報量で、同じ確率で起こる 2 つの事象の内どちらが起こったかを知ることができます。

例として、硬貨投げで表が出るか裏が出るかという問題を考えると、表が出る確率は $\dfrac{1}{2}$ ですから、表が出たことを知るための情報量は $-\log_2 \dfrac{1}{2} = 1$ ビットになります。逆に考えれば、1 ビットあれば表を 1・裏を 0 とすることで、どちらが起こったのか必要な情報を得ることができます。もう一つの例としてサイコロの問題を考えてみましょう。1 が出たことを知るための情報量は $-\log_2 \dfrac{1}{6} = 2.58\cdots$ となり、3 ビット必要になります。逆に、1 を 001, 2 を 010, 3 を 011, 4 を 100, 5 を 101, 6 を 110 にそれぞれ対応させることによって、必要な情報が得られます。上記の 2 つの例から、n ビットでは 2^n 個の情報を得ることができ、起こる確率が小さいほど情報量は大きくなることが判ります。

[6] 当時は、「情報」と「状報」が使われていたが、明治 30 年以降には「情報」統一された。また、「情報」という言葉が一般人の目にふれるようになったのは日清戦争 (明治 27 年) の時で、戦況を伝える新聞記事に使用された。

その他の情報量を表す単位として、8 ビット (2 進数 8 桁) をひとまとめにして扱う**バイト** (byte) という単位があります (8bit=1byte)。さらに、バイトを基準として 8 の倍数 (8, 16, 24, 32, 40, 48, 56, 64, ⋯) のビットをひとまとめにして扱う**ワード** (word) と呼ばれる単位があります。ワードは、コンピュータ内部でプログラムやデータを扱う基本単位で、コンピュータの進化と共に大きな値を取るようになり、現在では 32 ビットまたは 64 ビットが 1 ワードとなっています。なお、図 2.3 のように 1 ワードを 2 バイト (16 ビット) で表すとき、左から順に各桁を「第 15 ビット (15 ビット目)」, 「第 14 ビット (14 ビット目)」, ⋯, 「第 0 ビット[7] (0 ビット目)」と呼び、特に、左端の第 15 ビットを**最上位ビット**、右端の第 0 ビットを**最下位ビット**と呼びます。また、最上位ビットから 8 ビット目までを上位 8 ビット、最下位ビットから 3 ビット目までを下位 4 ビットのように呼びます。

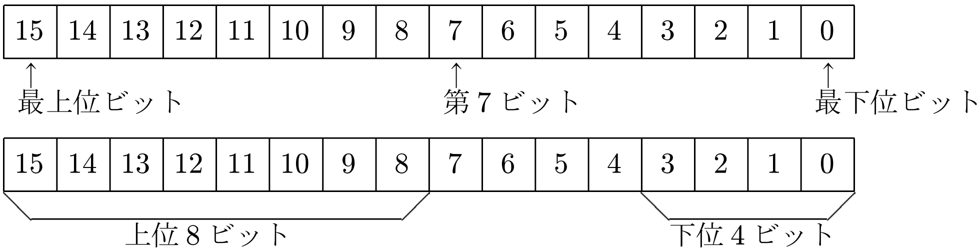

図 2.3: 1 ワードを 2 バイトとした場合

上記で述べたように、コンピュータは 1 ワードを基本単位として情報を扱います。では、逆に、この 1 ワードで表すことのできる情報を自然に整数に対応させると**いくつ**から**いくつ**までの数を表すことが出来るのでしょうか。例えば、1 ワードが 8 ビットの場合、

$$
\begin{array}{ccc}
\texttt{00000000(2)} & \longleftrightarrow & \texttt{0} \\
\texttt{00000001(2)} & \longleftrightarrow & \texttt{1} \\
\texttt{00000010(2)} & \longleftrightarrow & \texttt{2} \\
\vdots & \vdots & \vdots \\
\texttt{11111110(2)} & \longleftrightarrow & \texttt{254} \\
\texttt{11111111(2)} & \longleftrightarrow & \texttt{255}
\end{array}
$$

のように自然に対応させることで 0 から 255 $(= 2^8 - 1)$ までの整数を表わすことができます。その他に、情報科学の分野でよく用いられるビット数で表現可能な整数の範囲を表 2.7 に挙げておきます。

8 ビット	0 から 255 $(= 2^8 - 1)$
16 ビット	0 から 65535 $(= 2^{16} - 1)$
24 ビット	0 から 16777215 $(= 2^{24} - 1)$
32 ビット	0 から 4294967295 $(= 2^{32} - 1)$
64 ビット	0 から 18446744073709551615 $(= 2^{64} - 1)$

表 2.7: 1 ワードで表現可能な整数の範囲

[7]第 16 ビット, 第 15 ビット, ⋯, 第 1 ビットと 1 から数える場合もあるが、本テキストでは 0 から数える。

一般的な場合には、組み合わせの数を考えることで、r 進法 k 桁で表現可能な整数の範囲は、

$$0 \text{ から } r^k - 1 \quad (0 \text{ から始まるため } 1 \text{ を引く})$$

までとなります。

$\boxed{\textbf{例題 1}}$　1 から 1000 までの数字の内、いずれの数字であるかを知るために必要な情報量を求めなさい。

解答例　$2^9(= 512) < 1000 < 2^{10}(= 1024)$ より、$\dfrac{1}{2^{10}} < \dfrac{1}{1000} < \dfrac{1}{2^9}$ であるから、情報量は

$$-\log_2 \frac{1}{2^{10}} > -\log_2 \frac{1}{1000} > -\log_2 \frac{1}{2^9}$$

$$\iff \quad 9 = \log_2 2^9 < \log_2 1000 < \log_2 2^{10} = 10$$

となる。従って、1 から 1000 までの数字の内、いずれの数字であるかを知るためには 10 ビット必要である ($\log_2 10 = 3.321\cdots$ を知っていれば、情報量は

$$-\log_2 \frac{1}{1000} = \log_2 10^3 = 3\log_2 10 = 3 \times 3.321\cdots = 9.96\cdots$$

より、10 ビット必要であることが直ちに得られる)。

問題 1　1 から 100 までの数字の内、いずれの数字であるかを知るために必要な情報量を求めなさい。

問題 2　アルファベット (A〜Z) の内、いずれのアルファベットであるかを知るために必要な情報量を求めなさい。

問題 3　トランプ 52 $(= 13 \times 4)$ 枚の内、スペード・ダイア・ハート・クローバのいずれであるかを知るために必要な情報量を求めなさい。

2.5　補数表現とイクセス表現

ある基準となる決まった数から別の数を引いた数を**補数** (complement) と呼び、このような相対的な数の表現方法を**補数表現**と呼びます。情報科学の分野では、**負の整数を正の整数で表現する方法**として用いられます。具体的には、負の r 進数 $-a$ ($a > 0$) の補数は、ある決まった正の整数 r^k に $-a$ を足すことで得られます (これを「**r 進数 k 桁で補数表現する**」という)。

例えば、ある決まった正の整数を 10 としたとき (10 進数 1 桁で補数表現するとき)、10 進数 -3 の補数は、

$$10 + (-3) = 7$$

となります。言い換えると、10 進数 1 桁で補数表現された世界の 7 は -3 を意味します。同様に、

$$\begin{aligned}
&10\text{ 進数 } -1 \text{ の補数は } 10 + (-1) = 9,\\
&10\text{ 進数 } -2 \text{ の補数は } 10 + (-2) = 8,\\
&10\text{ 進数 } -3 \text{ の補数は } 10 + (-3) = 7,\\
&10\text{ 進数 } -4 \text{ の補数は } 10 + (-4) = 6,\\
&10\text{ 進数 } -5 \text{ の補数は } 10 + (-5) = 5
\end{aligned}$$

となり、10 進数 1 桁の補数表現によって -5 から -1 までの負の整数を正の整数に対応させて表現することが出来ます。すなわち、10 進数 1 桁で表現可能な正の整数 0 から 9 の内、半分が負の整数に割り当てられ、残り半分が**そのまま 0 と 1 から 4 までの正の整数に割り当てられます** (**補数表現の場合、正の整数も同じ k 桁で表示すること**)。逆に、補数表現された負の整数を元 (通常) の負の整数に戻すには、補数表現された負の整数 (補数) からある決まった正の整数を引きます。例えば、10 進数 1 桁で補数表現された負の整数 8 を元 (通常) の負の整数に戻すと、

$$8 - 10 = -2$$

となります ($\because -2 = (10 - 10) - 2 = (10 - 2) - 10 = 8 - 10$)。

ここで、補数表現を使った演算について考察して観ましょう。結論から言うと、補数表現を用いると引き算を足し算として扱うことが出来ます[8]。例えば、$4 - 3$ を 10 進数 1 桁で補数表現し直して計算を行うと、

$$(4 - 3 = 4 + (-3) =)\ 4 + 7 = \underline{1}1$$

が得られます。補数表現を使った演算では、決まった桁数 (ここでは 1 桁) だけを扱うため、**下線部を無視する**ことで 1 という正しい答えが得られます。その他にも、

$$\begin{aligned}
3 - 1\ (= 3 + (-1)) &\quad \text{を補数表現すると } 3 + 9 = \underline{1}2\,(= 2),\\
1 - 1\ (= 1 + (-1)) &\quad \text{を補数表現すると } 1 + 9 = \underline{1}0\,(= 0),\\
1 - 3\ (= 1 + (-3)) &\quad \text{を補数表現すると } 1 + 7 = \ \ 8\,(= -2),\\
-2 - 2\ (= (-2) + (-2)) &\quad \text{を補数表現すると } 8 + 8 = \underline{1}6\,(= -4)
\end{aligned}$$

などが成り立ち、補数表現を使っても正しく演算できることがわかります。

[8] 負の整数を正の整数として、引き算を足し算として扱うことが出来ると、演算装置に減算回路が不必要 (加算回路だけで十分) となり、演算回路を単純にすることが出来ます (詳しくは 4 章で)。

理解を深めるために、他の例について観てみましょう。10進数3桁で補数表現した場合 (ある決まった正の数を $1000 = 10^3$ とする)、表 2.8 のように -500 から 499 までの整数を表現することができます。正の整数と負の整数の判別は、第3桁が 0, 1, 2, 3, 4 であれば正の整数、5, 6, 7, 8, 9 であれば負の整数となります。同様に、2進数8桁で補数表現した場合 (ある決まった正の数を $512 = 2^8$ とする)、表 2.9 (付録 A.2 参照) のように -256 から 255 までの整数を表現することができます。なお、正の整数と負の整数の判別については、10進数に比べると容易に判断することができ、**最上位ビットが0であれば正の整数、1であれば負の整数**となります。

実際の数	補数表現
-500	500
-499	501
-498	502
⋮	⋮
-3	997
-2	998
-1	999
0	000
1	001
2	002
3	003
⋮	⋮
497	497
498	498
499	499

表 2.8: 10進数3桁の補数表現

実際の数	補数表現(2)
-128	10000000
-127	10000001
-126	10000010
⋮	⋮
-3	11111101
-2	11111110
-1	11111111
0	00000000
1	00000001
2	00000010
3	00000011
⋮	⋮
125	01111101
126	01111110
127	01111111

表 2.9: 2進数8桁の補数表現

これまでのまとめとして、負の r 進数 $-a$ $(0 < a \leq r^k/2)$ から r 進数 k 桁の補数表現への変換は

$$r^k + (-a) = r^k - a$$

によって得られ、逆に、r 進数 k 桁で補数表現された補数 b $(= r^k - a)$ から元 (通常) の負の r 進数への変換は

$$b - r^k = -(r^k - b)$$

によって得られます ($\because b - r^k = (r^k - a) - r^k = -a$)。まとめると、$r$ 進数 k 桁で補数表現された補数0から $r^k - 1$ の内 (**補数表現の場合、正の整数も同じ k 桁で表示すること**)、

$$\begin{cases} \dfrac{r^k}{2}\ \text{から}\ r^k - 1\ \text{を負の整数} & \left(\text{元の整数}\ -\dfrac{r^k}{2}\text{から}\ -1\ \text{に対応}\right), \\ \quad 0 \qquad\quad \text{を} \qquad\quad 0, & \\ 1\ \text{から}\ \dfrac{r^k}{2} - 1\ \text{を正の整数} & \left(\text{元の整数}\ 1\ \text{から}\dfrac{r^k}{2} - 1\ \text{に対応}\right) \end{cases}$$

のように半分が負の整数に対応します (注意 : 演算を行う際は、上記の範囲内で行う必要がある)。

実は、補数表現には 2 種類の表現方法があって、r 進数 k 桁で補数表現するとき、これまで学習してきた **r の補数**と、ある決まった正の整数を r^k-1 とした **$r-1$ の補数**があります (本テキストでは、「補数を求めなさい」のように r の補数であるか $r-1$ の補数であるかが省略されている場合、基本的に r の補数を指します)。r の補数と $r-1$ の補数の間には、負の r 進数 $-a$ $(a>0)$ に対して r の補数が $r^k+(-a)$ によって得られることより、

$$\Big(r^k+(-a)\Big)=r^k+(-1+1)+(-a)=\Big((r^k-1)+(-a)\Big)+1$$

のように変形することで、

$$「r \text{ の補数}」=「r-1 \text{ の補数}」+1$$

という関係が成り立ちます。まとめると、r 進数 k 桁で補数表現された 1 の補数 0 から r^k-1 の内 (**補数表現の場合、正の整数も同じ k 桁で表示すること**)、

$$\begin{cases} \dfrac{r^k}{2} \text{ から } r^k-2 \text{ を負の整数} & \left(\text{元の整数 } -\dfrac{r^k}{2}+1 \text{ から } -1 \text{ に対応}\right), \\ \quad r^k-1 \quad \text{を} \quad -0, & \\ \quad\quad 0 \quad\quad \text{を} \quad +0, & \\ 1 \text{ から } \dfrac{r^k}{2}-1 \text{ を正の整数} & \left(\text{元の整数 } 1 \text{ から } \dfrac{r^k}{2}-1 \text{ に対応}\right) \end{cases}$$

のように約半分が負の整数に対応します (注意 : $r-1$ の補数には 0 の表現方法が 2 種類存在し、「+0 (プラスのゼロ)」及び「−0 (マイナスのゼロ)」と呼ばれる)。

例として、10 進数 3 桁で補数表現すると (ある決まった正の整数を 10^3-1 $(=999)$ としたとき)、10 進数 -10 の **9 の補数**は、

$$(10^3-1)+(-10)=999-10=989$$

となります。なお、

```
    999
 -   10
 ------
    989
```

のように記述すると、各桁ごとに引き算すればよく、繰り下がりを気にすることなく容易に計算することが出来ます。もちろん、9 の補数に 1 を足せば **10 の補数** 990 $(=989+1)$ になることは明らかです。**$r-1$ の補数を求めてから 1 を足して r の補数を求めると、計算が楽な上、計算ミスが少なくなります。**

コンピュータ上で補数を扱う際に、**1 の補数** (付録 A.2 参照) を考えることは非常に重要な意味を持ちます。例えば、2 進数 8 桁で補数表現するとき、$-1011(2)$ の **2 の補数**は

$$100000000(2) + (-1011(2))$$

を計算しても求められますが、

$$100000000(2) + (-1011(2)) + (-1(2) + 1(2)) = \underline{11111111(2) + (-1011(2))} + 1(2)$$

のように下線部の 1 の補数を計算してから、1 を足しても求められます。ここで、1 の補数の計算

$$\begin{array}{rr} & 11111111(2) \\ - & 00001011(2) \\ \hline & 11110100(2) \end{array}$$

を注意深く観察すると、11110100(2) は 00001011(2) の 0 と 1 をひっくり返す操作を行うだけで得られることがわかります。3 章で詳しく述べますが、1 の補数の計算は、論理演算の否定と全く同じ働きをしています[9] (コンピュータにとっては非常に有利なことである)。

例題 1　10 進数 4 桁で補数表現するとき、$-123(10)$ について 10 の補数および 9 の補数を求めなさい。

解答例　10 の補数　10000−123=9877(10)
　　　　9 の補数　9999−123=9876(10)

例題 2　2 進数 8 桁で補数表現するとき、$-1010100(2)$ について 2 の補数および 1 の補数を求めなさい。

解答例　2 の補数　100000000−1010100=10101100(2)
　　　　1 の補数　11111111−1010100=10101011(2)

例題 3　2 進数 8 桁で補数表現された 2 の補数 10101001(2) を元の 2 進数に直しなさい。

解答例　10101001−100000000= −(100000000−10101001)= −1010111(2)

例題 4　2 進数 8 桁で補数表現された 1 の補数 10101001(2) を元の 2 進数に直しなさい。

解答例　10101001−11111111= −(11111111−10101001)= −1010110(2)

[9] 最終的には、四則演算を論理演算で再構築し、論理回路で実現します。

補数表現の類似として、**イクセス表現**[10] (excess expression) と呼ばれる数の表現方法があります。イクセス表現は、全ての整数に**バイアス** (bias) と呼ばれるある決まった正の整数を一様に足して数を表現します (イクセス表現もゼロと正の整数に対応させて数を表現します)。この表現方法の利点は、**正の数・ゼロ・負の数に関係なく元の数の大小関係を容易に判別できるという利点を持っています**。以下、**r 進数 k 桁でイクセス表現する**場合、話を簡単にするためにバイアス (ある決まった正の整数) を $\dfrac{r^k}{2}$ として話を進めていきます[11]。r 進数 a を r 進数 k 桁でイクセス表現するには、

$$a + \frac{r^k}{2}$$

を計算します。ただし、イクセス表現された数はゼロと正の整数に対応させることから、a の範囲は

$$-\frac{r^k}{2} \leq a \leq \frac{r^k}{2} - 1$$

となり、イクセス表現された a の範囲は

$$0 \leq a \leq r^k - 1$$

となります。逆に、r 進数 k 桁でイクセス表現された数 b $(= a + \dfrac{r^k}{2})$ から元の整数への変換は

$$b - \frac{r^k}{2}$$

によって得られます。まとめると、バイアスが $\dfrac{r^k}{2}$ のとき、r 進数 k 桁でイクセス表現された数 0 から $r^k - 1$ の内、

$$\begin{cases} 0 \quad \text{から}\ \dfrac{r^k}{2} - 1\ \text{を負の整数} & \left(\text{元の整数}\ -\dfrac{r^k}{2}\text{から}\ -1\ \text{に対応}\right), \\ \quad \dfrac{r^k}{2} \quad\quad \text{を} \quad\quad 0, & \\ \dfrac{r^k}{2} + 1\ \text{から}\ r^k - 1\ \text{を正の整数} & \left(\text{元の整数}\ 1\ \text{から}\ \dfrac{r^k}{2} - 1\ \text{に対応}\right) \end{cases}$$

のように半分が負の整数に対応します。

例として、バイアスを $\dfrac{10^3}{2}$ $(= 500)$ とした 10 進数 3 桁のイクセス表現を表 2.10 にバイアスを $\dfrac{2^8}{2}$ $(= 128)$ とした 2 進数 8 桁のイクセス表現を表 2.11 に挙げておきます (付録 A.2 は、バイアスを $\dfrac{2^8}{2} - 1$ $(= 127)$ で算出した値で、単精度 IEEE754 形式に対応しています。)。

[10] ある決まった正の数を一様に足すことで数を底上げして表示することから、「下駄履き表現」とも呼ばれます。この表示方法は 2.7 節の浮動小数点表示の指数部を表示するのに用いられます。

[11] 2.7 節の浮動小数点数 (IEEE754 形式) で用いられるイクセス表現のバイアスは $\dfrac{r^k}{2} - 1$ で算出します。

実際の数	補数表現	イクセス表現
-500	500	000
-499	501	001
-498	502	002
⋮	⋮	⋮
-3	997	497
-2	998	498
-1	999	499
0	000	500
1	001	501
2	002	502
3	003	503
⋮	⋮	⋮
497	497	997
498	498	998
499	499	999

表 2.10: 10 進数 3 桁のイクセス表現

実際の数	補数表現(2)	イクセス表現(2)
-128	10000000	00000000
-127	10000001	00000001
-126	10000010	00000010
⋮	⋮	⋮
-3	11111101	01111101
-2	11111110	01111110
-1	11111111	01111111
0	00000000	10000000
1	00000001	10000001
2	00000010	10000010
3	00000011	10000011
⋮	⋮	⋮
125	01111101	11111101
126	01111110	11111110
127	01111111	11111111

表 2.11: 2 進数 8 桁のイクセス表現

例題 5　123(10) を 10 進数 4 桁でイクセス表現しなさい。ただし、バイアスは $\frac{10^4}{2}$ $(= 5000)$ とする。

解答例　$123(10) + \frac{10^4}{2} = 123(10) + 5000(10) = 5123(10)$

例題 6　$-$123(10) を 10 進数 4 桁でイクセス表現しなさい。ただし、バイアスは $\frac{10^4}{2}$ $(= 5000)$ とする。

解答例　$-123(10) + \frac{10^4}{2} = -123(10) + 5000(10) = 4877(10)$

例題 7　101001(2) を 2 進数 8 桁でイクセス表現しなさい。ただし、バイアスは $\frac{2^8}{2}$ $(= 128)$ とする。

解答例　$101001(2) + \frac{2^8}{2} = 00101001(2) + 10000000(2) = 10101001(2)$

例題 8　$-$101001(2) を 2 進数 8 桁でイクセス表現しなさい。ただし、バイアスは $\frac{2^8}{2} - 1$ $(= 127)$ とする。

解答例　$-101001(2) + \left(\frac{2^8}{2} - 1\right) = -101001(2) + 1111111(2) = 01010110(2)$

例題 9　2 進数 8 桁でイクセス表現された数 10101010(2) を元の 2 進数に戻しなさい。ただし、バイアスは $\frac{2^8}{2}$ $(= 128)$ とする。

解答例　10101010(2) $- \frac{2^8}{2} =$ 10101010(2) $-$ 10000000(2) $=$ 101010(2)

問題 1　10 進数 8 桁で補数表現するとき、$-$123(10) を 9 の補数と 10 の補数で表しなさい。

問題 2　4 進数 4 桁で補数表現するとき、$-$123(4) を 3 の補数と 4 の補数で表しなさい。

問題 3　2 進数 16 桁で補数表現するとき、$-$1000(10) を 2 進数に直し、2 の補数で表しなさい。

問題 4　2 進数 16 桁で補数表現された 2 の補数 1111000011110000(2) を元の 2 進数に直し、10 進数で答えなさい。

問題 5　2 進数 16 桁で補数表現された 1 の補数 1111000011110000(2) を元の 2 進数に直し、10 進数で答えなさい。

問題 6　123(4) を 4 進数 4 桁でイクセス表現しなさい。ただし、バイアスは $\frac{4^4}{2}$ とする。

問題 7　$-$10111(2) を 2 進数 8 桁でイクセス表現しなさい。ただし、バイアスは $\frac{2^8}{2} - 1$ とする。

問題 8　2 進数 8 桁でイクセス表現された数 01010101(2) を元の 2 進数に戻し、10 進数で答えなさい。ただし、バイアスは $\frac{2^8}{2} - 1$ とする。

問題 9　10 進数 4 桁でイクセス表現された数 1234(4) を元の 10 進数に戻しなさい。ただし、バイアスは $\frac{10^4}{2}$ とする。