

n 進小数と符号と極限

ふじ い じゅん いち
藤 井 淳 一

(大阪教育大学 教育協働学科 理数情報講座)

(2020 年 3 月 14 日 受付)

概要： n 進小数を巡る話題として、教育的な意味での位置づけをして、有効活用の場面で「算術符号」を例示してみた。無限小数の概念を巡って、圏論的な 2 つの極限と解析的な極限の差異についても俯瞰してみたい。

検索語：小数の構造、 n 進小数、瞬時符号、算術符号、圏論的極限

はじめに

小数というのが歴史に登場したのは、意外なことに「多項式」「対数」の登場時期とさほど変わらない。実際、対数は一応 Napier の 1614 年の著作「Mirifici Logarithmorum Canonis Descriptio (奇蹟の対数法則の仕組み)」が嚆矢とされている [5] が、対数自体は 1595 年あたりから研究していたらしい。多項式は、「指数」を意識し始め、虚数概念を確立させた Bombelli の「L'Algebra」(1572 年)あたりで、10 進小数はまさに「De Thiende (10 進法、正確には「10 分の 1 法」かもしれない)」を 1585 年に著した Stevin あたりかと思われる。彼は同年に「L' Arithmetique (算術)」を著して多項式から代数学の考察をしている。

実は小数点の発明者は、Napier ではないかともいわれていて、Stevin 以前に小数についても研究していたようである。逆に Stevin も、Galileo に先んじて数学的自然科学を提唱し、応用数学の中で対数の研究をしていた。このように、時代的にも人物的にもこれらの概念の成立は密接にかかわっている。ただし、劉徽『九章算術』(263 年)に既に小数表記があったともいわれている。解釈によって異論はあるだろうが、ヨーロッパではほぼ同時期であることは間違いのないのではなかろうか。因みに私がよく引用する Bochner [2] では、「代数学が中世後期における商業の勃興の圧倒的な影響の下で、『算術』との双生児のような現象として生み出された」と表現している。

ただ、その現象の説明は、数学をやっている人なら、多項式

$$p(X) = c_0 + c_1X + c_2X^2 + \cdots + c_kX^k$$

に対し、 n 進構造は、 $X = \frac{1}{n}$, $c_k \in \{0, 1, \dots, n-1\}$ と対応させた (ただし、 a_0 のみは $n-1$ までの制限をかけない場合がある) $p(\frac{1}{n})$ そのものであり、その時の小数以下の桁数 k は、 n 進「整数」に変換した $p(n)$ を考えた場合、 k 桁の範囲が n^k 以上 $((n+1)^k$ 未満) なので

$$k = \max\{\ell \in \mathbb{N} \mid \ell \leq \log_n p(n)\}$$

となることから、なんとなく納得できるのではないだろうか。

[3]でも少し触れたが、特に n 進数を意識した小数の教育的取扱いについては、あまり見かけたことがない。大学生でも n 進整数の変換はできても、小数となるとすぐには対応できないようである。そこで上記のような認識の下で、小数を巡る話題として考えたことを記しておきたい。

I. 有限小数

小数が「有限か循環無限か」という区別は、あまり絶対的なものではないということを意識されたことはあるだろうか？ n 進数の観点からは、本来の「数」が持つ量的な性質において、特に有理数を扱うレベルでは有限・無限の区別に絶対性はない。循環小数が分数で表されることは、学校数学でもやったことがある方は多いのではないだろうか？そこで、10 進数の観点からは、(小数の整数部分は自由に取れることにして)「有理数=有限小数 \cup 循環小数」の disjoint union として扱われているだろう。しかし、ある循環小数が $\frac{m}{n}$ であるなら、 n 進数では、有限小数である： m が n 進表記で $(c_k \cdots c_1 c_0)_n$ とあらわされるとき、 m/n は、小数以下 1 桁の $(c_k \cdots c_1.c_0)_n$ になるだけである(後に変な方法で確認する)。

小数の計算は多項式の観点から見ても、各項でまず計算し、係数条件 $c_k \in \{0, 1, \dots, n-1\}$ に当てはまるように「繰り上げ」「繰り下げ」が行われることになる。このあたりの観点があまり意識されてないと、 n 進数(特に小数)の計算で戸惑うことになる。学生の計算の様子を見てみると、各次数の単項式による同類項のまとめは、各桁の計算そのものであり、多項式計算と n 進計算の違いところは繰り上がり繰り下がりのみだという認識が案外できていないように思える。

変換に戸惑いを見せた大学生でも、多項式との関連の話をすると比較的変換できるようになる。ただ、分数で適当に分配して n 進小数変換をしているようで、整数のように効率の良いアルゴリズムには気づいていないようである。そこで少し変換にふれておこう。次章のように小数以下の桁数 k が決まっている場合には、 n^k 倍して整数部分のみを n 進整数変換の適当なアルゴリズムで奥菜えばよい。簡単のため 1 未満の正の小数 ($a_0 = 0$) にして、述べると、小数以下の n 進の各桁の数は、「 n 倍の整数部分を拾って、小数部分について同じことを順次繰り返す」ことで得られることまでは、なかなか気づいていないようである。また、このように席をとって変換するアルゴリズムは、 n で割って行って、商として各桁の賞を求めているのと同様であるが、アルゴリズムとしてはずっとシンプルであり、「 n 進小数の変換の方が簡単」なのである。

この積・商 2 つのアルゴリズムの同等性に目を向けると、循環小数の話とつながる。小数部分 1 桁の n 進小数では、積のアルゴリズムを 1 回繰り返して終了する。それは商の方でも同じなのであるが、仮に小数以下 1 桁の n 進小数の唯一の桁の数が 1 の時； $(0.1)_n$ 、割り算で 1 少ない商 $n-1$ を立てても矛盾せず無限に続くため、 $(0.0(n-1))_n$ という循環小数になる。10 進数では、 $0.1 = 0.0\dot{9}$ ということである。逆に 10 進の循環小数 N で、循環部分が整数で表したとき $M = d_1 \cdots d_m$ とする； $N = 0.\dot{d}_1 d_2 \cdots d_{m-1} \dot{d}_m$ 。これは、 M でくくると、

$$N = M \times \overbrace{0.00 \cdots 01}^{m \text{ 桁}}.$$

したがって、同じ桁だけの 9 が並ぶ数をかけると、

$$\overbrace{9 \cdots 9}^{m \text{ 桁}} N = M \times \overbrace{0.99 \cdots 99}^{m \text{ 桁}} = M \times 0.\dot{9} = M$$

となって、すくなくとも $n = \overbrace{9 \cdots 9}^m$ 進小数としては有限である。10 進と n 進間のみの変換例を挙げたが、どの進数間でも同等である。

ところで、最初の国定教科書「黒表紙」を作成した藤沢利喜太郎がいわゆる「尾加法」の筆算を導入したことはご存じだろうか？小さい桁から計算することで、合理的に繰り上げの煩雑さを整理した方法である。これ自体は立派な「発明」ではあり、日本人の計算能力を向上させた点からも有効性を否定するものではないが、これによって「暗算」の有効性を否定されたのであれば問題が生じる。日本人的にきっちと1円の間違いもなしに計算させるという目的からは外れているが、「暗算」は「概算」という観点からは有効であり、それは基本「頭加法」である。算盤をやったことのある人ならわかると思うが、通常の計算も含めてすべて「頭から」の計算である。「商業算術」という位置づけは、当時から問題があったのではないだろうか？

せっかくフランスから寺尾寿が持ち帰った優れた「理論算術」を否定し、「商業算術」を標榜したことは今となってはどのように評価すべきだろうか？実際使われていた「商業の算術」は昔は算盤、今は電卓・コンピュータである。お題目は現場とは、完全にずれていたのである。日常生活で、筆算を使うことはあまりないと思われるが、概算としての暗算は使われているのではないだろうか？日本の算数数学教育の歴史を振り返るたびに、「理論算術」「緑表紙の数理思想」を有効に使えなかったことは、つくづく残念に思う。

II. 算術符号

さて、 n 進小数の応用例として、すぐに思いつくのは算術符号である。情報科学の分野に属していながら、あまり通常の古典的な情報理論関連について触れてこなかったが、ここでは小数を使った Shannon-Fano の算術符号と呼ばれる古典そのものを例として情報源符号について触れてみたい。その前に一般的な情報符号についての解説をしておこう。

上記で「情報源符号」という言い方をしたが、その対義語は「通信路符号」である。符号化には2つの相反する目的がある：

- (1) 情報をできるだけ圧縮すること
- (2) 符号化は雑音で乱されてもある程度修復可能であること

である。前者の情報圧縮に対応するのが「情報源符号化」であり、後者の誤り訂正可能性を持たせるものが「通信路符号化」である。情報をぎりぎりまで圧縮すると、(経年劣化や通信途中の外的な雑音等で) ちょっとでも乱されると使い物にならなくなる。情報は永久不変ではない。1bit の間違いで、爆発事故が起きることはよく聞く話である。そんなことにならないよう、実用的にはこれらを目的に応じてバランスよく組み合わせる必要がある。実際 FAX では、ランレングス符号化やハフマン符号化で情報圧縮し、巡回符号化で訂正可能にしている。また、前者は圧縮に明確な限界があるため、ハフマン符号などの最適解がすでに出ているが、後者は訂正可能な能力を挙げれば上げるほど情報は膨れ上がるため、実用性との兼ね合いになり、用途の違いによっても様々な「誤り訂正符号」がある。現在「符号理論」と呼ばれ、研究が続けられているのはこちらの方である。

したがって、「情報源符号化」は古典となっているが、この符号化にもいくつか種類がある。符号の長さで「固定長」「可変長」があり、前者はどの部分にどんな情報があるかの「フォーマット」が決められているため、扱いは簡単であるが、すべて同じ形にするために無駄が多い。情報圧縮の観点からは「可変長」でないと効果がないだろう。さらに性質として、機械的な言語であるから「多義性」は排除し「一意符号」であるべきだろう。単語としての一意性のみでは不十分である。文字列として連なったときにも一意的な解釈ができないといけない。つまり、「うらにわをかいた」という多義的な文になってはダメなので「可変長」の場合は「単語が一意的に区切れること」が大切である。もちろん自然言語のような句読点やカンマ・ピリオドによる区切りの区別という手もあるだろう。しかし、究極の情報圧縮を目指す場合にはそれすら「無駄」に属する。さらに「夢落ち」という厄介な問題がある。述語が最後に出てくる日本語のような構造だと、「.. というのは夢だった」とすべてを最後の句でひっくり返すことができるが、このような言語構造は望ましくない。すべてその時点でタイムラグなしに情報が確定する「瞬時符号」が求められるものである。実は任意長の一意符号は常に「同じ単語長の」瞬時符号に直せるという McMillan の定理がある（つまりは設計はあっても作り方が下手ということである）。

瞬時符号となる同値条件は「木符号として葉のみに符号語を割り当てること」である。「木符号」とはグラフ理論的な「木」の各枝に一つずつアルファベットを割り当て、上記の場合だと「葉」にたどり着くまでのアルファベットの並びを「符号」の単語（符号語）に採用することである。これは「頭語条件」とも呼ばれる「同じアルファベット並びで、途中で終わるものと続くものがあってはならない」という瞬時符号条件に合致している。したがって情報源符号の一つの決定版「ハフマン符号」は木符号として説明されるのが普通である。

さて、その瞬時符号として古典的な **Shannon-Fano 符号** の話がやっとできるところまできた。これは2進小数を使った2進符号であるが、一気に定義すれば、「出現確率（の和）に応じて、2進小数展開したときに採用する桁数 ℓ_1 を元の確率に応じて決めたもの」であるが、詳細を述べる：符号化されるべき単語 u_k の出現確率 $p(u_k)$ は降順になっているとする。 $q_1 = 0$ とし、 $j \geq 2$ の加算確率を $q_j = \sum_{i=1}^{j-1} p(u_i)$ として、2進小数展開をする（この加算確率によって瞬時符号条件を満たすように設計されている）。そのとき、Shannon の情報源符号化定理¹⁾の基礎条件：

$$-\log_2 p(u_i) \leq \ell_i < -\log_2 p(u_i) + 1 \quad \left(H(U) \equiv - \sum_i p(u_i) \log_2 p(u_i) \leq \frac{\text{平均符号長}}{\ell} < H(U) + 1 \text{ に対応} \right)$$

を満たすように小数以下で採用する桁数 ℓ_i を決める（つまり $-\log_2 p(u_i)$ の切り上げ）。例示してみよう：

p	ℓ	q	$(q)_2$	符号
6/16	2	0	0.0000	00
4/16	2	6/16	0.0110	01
3/16	3	10/16	0.1010	101
2/16	3	13/16	0.1101	110
1/16	4	15/16	0.1111	1111

（下線部分是对数の値がちょうど整数になるところである）。これでちゃんと瞬時符号ができている。確率を加算することで、最終的に1が続く形となり、単語の終わりが00, 01, 10, 11（最後のだけは、単

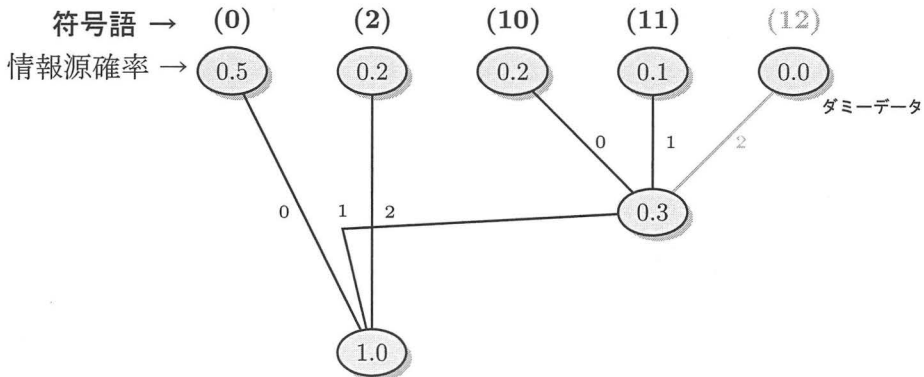
¹⁾本格的なものは、いくらでもエントロピーに平均符号長を近づけられるということであるが、これはアルファベットを組にして新たにアルファベット化すれば実現できる。

語長 4 以下で判断) と明確に句切れもわかる。しかも平均符号長は (ハフマン符号のように) 最短ではないが、エントロピーと 1 以内の誤差である。

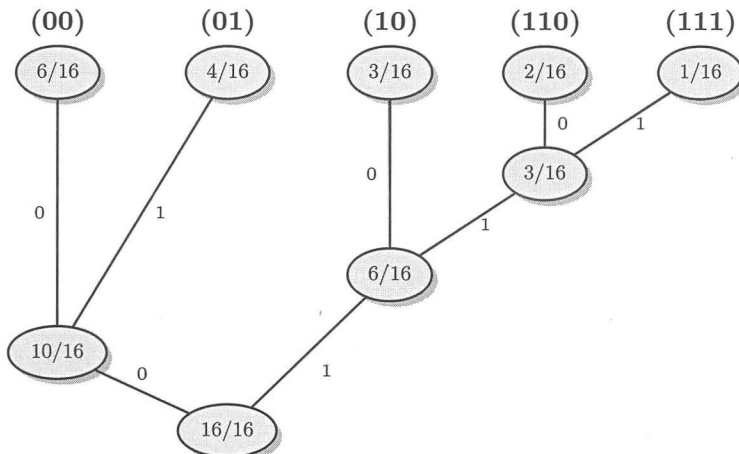
ちなみに、 n 進小数で同様の n 元 $(0, 1, \dots, n-1)$ 符号が得られる。情報量の不等式部分は、 n が底となる対数に代わるだけである。ほとんど見かけないので、上記の情報源確率で $n = 3$ の瞬時符号構成例を挙げておこう：

p	ℓ	q	$(q)_3$	符号
6/16	1	0	0.000	0
4/16	2	6/16	0.101	10
3/16	2	10/16	0.121	12
2/16	2	13/16	0.210	21
1/16	2	15/16	0.221	22

また今回の話とは外れるが、最短のハフマン符号は、逆に確率降順を基礎として、 n 個のアルファベットなら、後ろ n 個を仮想の葉としてまとめ、同じ手順を繰り返して「ハフマン木」を完成させて木符号を作るものである。ただし、 $n > 2$ の場合には最後 n 個をまとめてとはいかないので、準備として確率 0 のダミーを加えておかねばならない。例として 0.5, 0.2, 0.2, 0.1 の確率を持つ情報源の 3 元 (アルファベット 0, 1, 2) のハフマン符号の木の構成を挙げておく：



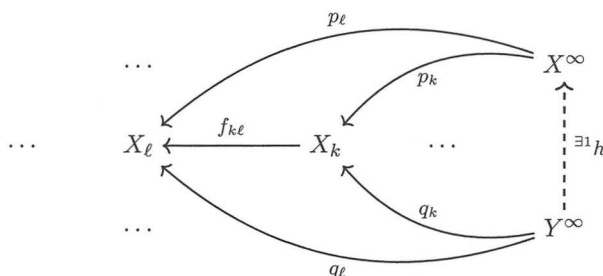
因みに Shannon-Fano の例を 2 元ハフマン符号化すると、明らかに短い符号ができる：



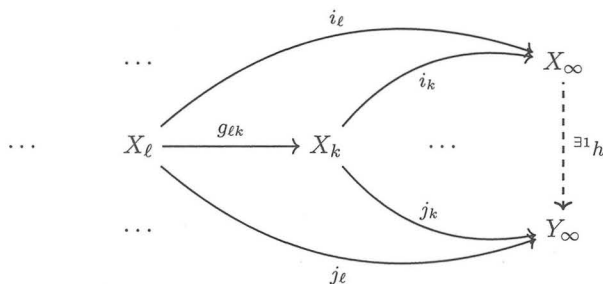
III. 圏論的極限と無限小数

桁数をそろえた小数（整数部分フリーにする）は、 n 進数で小数以下 k 桁の場合、集合として $\frac{1}{n^k} \mathbb{Z}$ と書けるが、整数倍で閉じているので、いわゆる \mathbb{Z} -加群（モジュール）である。 \mathbb{Z} -加群の圏（対象は \mathbb{Z} -加群、射は \mathbb{Z} -加群準同型）を想定して、極限の関係を考察しよう。

ここで少しだけ圏論的な極限の概念に触れる。上記のように（自然数とは限らない）有向インデックス k による対象 X_k があるとき、極限には top-down, bottom-up の 2 通りがある。前者 X^∞ は「直積」と関連し、射影極限、逆極限と呼ばれ $\varprojlim X_k$ と書かれる。正確な定義は、すべての k, ℓ について、 $f_{k\ell}: X_k \rightarrow X_\ell$ ($k \geq \ell$ の時、大きい集合を小さい集合に切り落として制限する射（射影）を想定) と、 X から X_k への射影 $p_k: X^\infty \rightarrow X_k$ があり、別の可能性 Y^∞ でも同様の $q_k: Y^\infty \rightarrow X_k$ があるとき、実は同類の Y^∞ があっても、すべて X^∞ が仕切っているかの如く、すべての合成が可換になる射 $h: Y^\infty \rightarrow X^\infty$ が一意的に存在するもので、以下の図式が可換なものと言い換えてもよい：



後者の方 X_∞ は直和と関連があり、上記と逆の関係になっていて、やはりこの方向での総元締めとなるよう $h: X^\infty \rightarrow Y^\infty$ が一意的に存在し、帰納極限、直極限などと呼ばれ $\varinjlim X_k$ と書かれる：



ここで、 $X_k = \frac{1}{n^k} \mathbb{Z}$ として、両極限を考えよう。 $\ell < k$ の場合には、 $X_\ell \subset X_k$ となっているので、 X_ℓ はそのまま X_k の中に写る（その写像が $g_{\ell k}$ である）。逆に小数以下 k 桁あるものを、少ない ℓ 桁までに切り落とす写像が $f_{k\ell}$ である。この時、切り落とす方は無限にあっても落とせるので、逆極限は $X^\infty \equiv \varprojlim X_k = \mathbb{R}$ になる。実際、すべての桁の小数を切り落とせるものはこれしかない。逆に、 n 進有限小数全体を \mathbb{Q}_n としたとき、すべての k で、 $X_k \subset \mathbb{Q}_n$ であるので、包含関係はこれを超えられず、 $X_\infty \equiv \varinjlim X_n = \mathbb{Q}_n$ になる。 n 進有限小数全体はそれなりに極限的にもまとまったものなのである。

さらに、 n も含めて、 $g_{(\ell,n)(k,n)} = g_{\ell k}$ とし、 m が n の倍数なら包含関係があるので、それを包含射 $g_{(\ell,n)(k,m)}$ としたときに、両有向インデックスを合わせた帰納極限を考えると、 $\varinjlim \frac{1}{n^k} \mathbb{Z} = \mathbb{Q}$ となる。い

ずれにせよ、無理数には帰納極限では届かない。つまり、無理数とは、どの n 進記法でも有限小数で表せない数の事であることがわかる。

圏論における極限概念は、すべての射を平等にみているためにこのような差異が生まれる。大きさの概念がないのである。数列の極限は大きさ（ノルム）の概念のもとの不平等な極限なので、bottom-up からでも無理数に到達できるという事情がある。平等性が必ずしもいい結果を生むとは限らない好例であろう。

ついでながら、極限概念として無理数を導入して「実数」を構成するのが自然だと思われる。そういう意味で「静的」な Dedekind 切断よりは、「動的」な Cauchy の実数論：

収束しそうな数列を「基本列」というが、有理数内では「完備」でないので収束先はないかもしれない。そこで、同じところに収束しそうな有理数列全体を同値類として、それ自身こそが「収束先」の実数と定めると、収束先がある場合もない場合も「すべて収束先が『実数』内にはある」＝「完備」となる。上記の同値関係による基本列全体のなす線形空間の商空間である。収束している「状態」を「収束先の実数」という「物」に転化させていることになる。

の方が自然であろう。

おわりに

本文と違って、通信路符号化の世界では決定版がないため、何らかの指針に従って研究を進めるのが普通である。その中に Goppa 符号を嚆矢とする「代数幾何符号」という難しい分野がある。難しいといったのは、代数幾何学といわれる分野のかなり高度な知識を使うからで、なかなか扱いが難しく、情報科学の分野に移ってからこの符号について院生に研究させることができたのは、一人だけだった。その方面の関連で [4] で少し触れたが、素数 p に関する「 p 進解析」と呼ばれる分野がある。これは最終章の話に密接にかかわるが、多分に数学的になりすぎるので本論では避けた。しかし、本論の用語は全く個人的なもので、記号や用語上正式なものと混同される危険性があるので、これ自体面白い話題ではあるし、最後に注意を込めて「既製品」に触れて置きたい。

素数 p について、 p 進整数環 \mathbb{Z}_p は、商空間の逆極限 $\varprojlim_n \frac{\mathbb{Z}}{p^n \mathbb{Z}}$ ($n \leq m$ に対し、射 $x + p^m \mathbb{Z} \mapsto x + p^n \mathbb{Z}$ を想定) であり、その分数を考えた「商体」を p 進体 \mathbb{Q}_p とするので、通常の実数体 \mathbb{Q} を含むものであることが大きく違う（濃度も \mathbb{R} と同じになる）。 p 進絶対値を

$$|\alpha|_p = \begin{cases} 0 & (\alpha = 0) \\ p^{-e} & (\alpha = p^e u) \end{cases}$$

とすると、(非アルキメデス) 付値と呼ばれる量になり、 $d_p(\alpha, \beta) = |\alpha - \beta|_p$ によって定まる p 進距離を導入すれば、 \mathbb{Q}_p は \mathbb{Q} の完備化（基本列の収束先を入れたもの）になっている。もちろん、この手の理論で指数や対数が出てくるのは当然である。

さらに大きく違う点は、 $1 = 0.9$ のようにならず、 \mathbb{Q}_p の元は一意的に桁展開できる点である。上記の距離が通常感覚と逆になっているので注意してほしい。 p^{-n} で p を大きくするとむしろ発散すること

になって一意性より 0 となるので、無限に（一意的に）桁展開できるのは大きい方である： $\sum_{j=-m}^{\infty} c_j p^j$. つ

まり大きい方に $p-1$ が無限に並ぶことはあり得て ($\alpha = \sum_{j=0}^{\infty} (p-1)p^j$)、その場合 1 を加えると順次上の桁に繰り上がってすべての桁が 0 になるので、 $\alpha = -1$ と解釈され、一意性が担保される（まるで無限のマジック！）。このように、上記展開ではマイナスの数はないように思えるが、ある桁以降の数がすべて $p-1$ になっている場合に対応し、マイナス記号なしですべての数が一意的に表されることになる。

このように興味深いもので実際数学上重要な概念であるが、通常感覚とはかなり違い現状の n 進数を（特に教育的な意味で）直接反映したものではないので、本論では避けたわけである（興味のある人のために、一つだけ参考に、多少通常の p 進小数を意識した講義録 [1] を挙げておこう）。

少し脱線したが、いずれにせよ、 n 進小数の計算程度は（整数より楽だし）高校でやってもいいように思われる。これをきっかけにして n 進小数にも目を向け、さらに符号理論にも興味を持っていただけたら幸いである。

参考文献

- [1] A.Baker, An introduction to p -adic numbers and p -adic analysis, Lecture Note, 2017.
<http://www.maths.gla.ac.uk/~ajb/dvi-ps/padicnotes.pdf>
- [2] S. ボホナー (村田 全 訳) : 科学史における数学, 1970 年, みすず書房.
- [3] 藤井淳一, 多項式環をめぐる, 数学教育研究, 15, 1985, 109–118.
- [4] 藤井淳一, Bost-Connes 代数と KMS 状態, 国際数理論理学協会会報, 76(2011), 4–13.
<http://www.jams.or.jp/kaiho/kaiho-76.pdf>
- [5] 成田収, 対数の誕生・成長・発展, 2005.
<http://www7a.biglobe.ne.jp/~watmas/dosukyo/circle-reports/logarithm.pdf>