計算理論

Thomas Zeugmann

北海道大学 大学院情報科学研究科
アルゴリズム研究室

http://www-alg.ist.hokudai.ac.jp/∼thomas/ToC/

第 8 回：$\mathcal{CF}$ と準同型写像

(日本語訳：湊 真一，大久保 好章)

代入 I

この講義では引き続き，文脈自由言語の有用な性質と特徴づけについて述べる．まず代入演算から見ていこう．

代入 I

この講義では引き続き，文脈自由言語の有用な性質と特徴づけについて述べる．まず代入演算から見ていこう．

定義 1

Σ と Δ を任意の 2 つの有限アルファベットとする．
写像 $\tau\colon \Sigma \longrightarrow \wp(\Delta^*)$ のことを**代入 (*substitution*)** と呼ぶ．我々は τ を写像 $\tau\colon \Sigma^* \longrightarrow \wp(\Delta^*)$ (つまり文字列) に拡張する．ただし次のように定義する．

(1) $\tau(\lambda) = \lambda$,

(2) すべての $w \in \Sigma^*$ と $x \in \Sigma$ に対して $\tau(wx) = \tau(w)\tau(x)$.

写像 τ は 次のようにして言語 $L \subseteq \Sigma^*$ に一般化される．

$$\tau(L) = \bigcup_{w \in L} \tau(w)$$

代入 II

つまり，代入演算は，Σ の各記号を Δ 上の言語に置き換える． 1 つの記号に対応する言語は，有限集合でも無限集合でもよい．

代入 II

つまり，代入演算は，Σ の各記号を Δ 上の言語に置き換える． 1つの記号に対応する言語は，有限集合でも無限集合でもよい．

例． 1

$\Sigma = \{0,1\}, \Delta = \{a,b\}$ とすると，$\tau(\lambda) = \lambda, \tau(0) = \{a\}$ および $\tau(1) = \{b\}^*$ により定義される写像 τ は代入である．

$\tau(010)$ を計算してみよう．定義より

$$\tau(010) = \tau(01)\tau(0) = \tau(0)\tau(1)\tau(0) = \{a\}\{b\}^*\{a\} = \underline{a}\langle\underline{b}\rangle\underline{a}\,,$$

ただし，最後の等式は正規表現の定義による．

代入 III

次に，代入演算に関する言語族 $\mathcal{L}$ の閉包が何を意味するかを定義しよう．ここで特別な考慮が必要である．最初にちょっと考えると，可能なすべての代入演算 τ について，条件 $\tau(L) \in \mathcal{L}$ を満たす必要があるように思われる．

しかし，この要求は強すぎる． なぜか?

代入 III

次に，代入演算に関する言語族 $\mathcal{L}$ の閉包が何を意味するかを定義しよう．ここで特別な考慮が必要である．最初にちょっと考えると，可能なすべての代入演算 τ について，条件 $\tau(L) \in \mathcal{L}$ を満たす必要があるように思われる．
しかし，この要求は強すぎる． なぜか?

$\Sigma = \{0, 1\}, \Delta = \{a, b\}$ かつ $\mathcal{L} = \mathcal{REG}$ としよう．さらに，$\tau(0) = L$ と仮定する．ただし L は Δ 上の任意の言語であって，再帰的に列挙可能だが非再帰的な言語であるとする．

代入 III

次に，代入演算に関する言語族 $\mathcal{L}$ の閉包が何を意味するかを定義しよう．ここで特別な考慮が必要である．最初にちょっと考えると，可能なすべての代入演算 τ について，条件 $\tau(L) \in \mathcal{L}$ を満たす必要があるように思われる．
しかし，この要求は強すぎる． なぜか?

$\Sigma = \{0,1\}, \Delta = \{a,b\}$ かつ $\mathcal{L} = \mathcal{REG}$ としよう．さらに，$\tau(0) = L$ と仮定する．ただし L は Δ 上の任意の言語であって，再帰的に列挙可能だが非再帰的な言語であるとする．

すると明らかに $\tau(\{0\}) = L$ もまた成り立つ．したがって $\tau(\{0\}) \notin \mathcal{REG}$ である．一方，$\{0\} \in \mathcal{REG}$ であるから, $\mathcal{REG}$ は代入演算に関して閉じていないことになる．同様にして $\mathcal{CF}$ も代入に関して閉じていないことが示される．

代入 IV

ここで考えないといけない点は，許される代入演算の集合を，Σ の要素から $\mathcal{L}$ に属する言語への写像だけに制限しなければいけないということである．したがって，以下の定義にたどり着く．

代入 IV

ここで考えないといけない点は，許される代入演算の集合を，Σ の要素から $\mathfrak{L}$ に属する言語への写像だけに制限しなければいけないということである．したがって，以下の定義にたどり着く．

定義 2

Σ を任意のアルファベットとし，$\mathfrak{L}$ を Σ 上の任意の言語族とする．もしもすべての写像 $\tau\colon \Sigma \longrightarrow \mathfrak{L}$ とすべての $L \in \mathfrak{L}$ に対して，$\tau(L) \in \mathfrak{L}$ が成り立つならば，$\mathfrak{L}$ は代入に関して閉じている *(closed under substitutions)* と言う．

準同型写像 I

定義 3

Σ と Δ を任意の 2 つの有限アルファベットとする．写像 $\varphi\colon \Sigma^* \longrightarrow \Delta^*$ は，

$$\text{すべての } v, w \in \Sigma^* \text{ に対して } \varphi(vw) = \varphi(v)\varphi(w)\,.$$

ならば**準同型写像 *(homomorphism)*** であると言う．さらに，

$$\text{すべての } w \in \Sigma^* \text{ に対して } \varphi(w) = \lambda \text{ ならば } w = \lambda\,.$$

ならば，φ は **λ-自由 *(λ-free)* 準同型写像**であるという．そして，$\varphi\colon \Sigma^* \longrightarrow \Delta^*$ が準同型写像ならば，**φ の逆写像 *(inverse)*** $\varphi^{-1}\colon \Delta^* \longrightarrow \wp(\Sigma^*)$ を定義できる．ただし各々の $s \in \Delta^*$ について

$$\varphi^{-1}(s) = \{w \mid w \in \Sigma^* \text{ かつ } \varphi(w) = s\}\,.$$

準同型写像 II

したがって，準同型写像は代入の特殊なケースである．

準同型写像 II

したがって，準同型写像は代入の特殊なケースである．
すなわち，準同型写像は，Σ の各記号を単一要素 (*Singleton*) の集合に対応づける．準同型写像の定義より明らかに，Σ の各記号に対する写像 φ を決めておけば十分である．なお，準同型写像を扱う場合，1 個の文字列だけを含む言語を表すために，その文字列そのものを記述することが一般的である.（すなわち，$\{s\}$ と書く代わりに，単に s と書く.）

準同型写像 II

したがって，準同型写像は代入の特殊なケースである．
すなわち，準同型写像は，Σ の各記号を単一要素 (*Singleton*) の集合に対応づける．準同型写像の定義より明らかに，Σ の各記号に対する写像 φ を決めておけば十分である．なお，準同型写像を扱う場合，1 個の文字列だけを含む言語を表すために，その文字列そのものを記述することが一般的である.（すなわち，$\{s\}$ と書く代わりに，単に s と書く.）

例． 2

$\Sigma = \{0,1\}$，$\Delta = \{a,b\}$ とする．$\varphi(0) = ab$ および $\varphi(1) = \lambda$ と定義された写像 $\varphi\colon \Sigma^* \longrightarrow \Delta^*$ は準同型であるが λ-自由な準同型ではない．
φ に 1100 を与えると，
$\varphi(1100) = \varphi(1)\varphi(1)\varphi(0)\varphi(0) = \lambda\lambda abab = abab$ が生成され，
言語 $\underline{1}\langle\underline{0}\rangle\underline{1}$ に対しては $\varphi(\underline{1}\langle\underline{0}\rangle\underline{1}) = \langle\underline{ab}\rangle$ となる．

注意点

今導入された概念の重要性を知るために，言語 $L = \{a^n b^n \mid n \in \mathbb{N}\}$ を考えてみよう．この言語は文脈自由であるから，直感的には $\{0^n 1^n \mid n \in \mathbb{N}\}$ もまた文脈自由である．なぜならば，我々は文法をたどりながら，出現するすべての a を 0 に，b を 1 に置き換えれば良いからである．

注意点

今導入された概念の重要性を知るために，言語 $L = \{a^n b^n \mid n \in \mathbb{N}\}$ を考えてみよう．この言語は文脈自由であるから，直感的には $\{0^n 1^n \mid n \in \mathbb{N}\}$ もまた文脈自由である．なぜならば，我々は文法をたどりながら，出現するすべての a を 0 に，b を 1 に置き換えれば良いからである．

この観察から，もしも出現するすべての a と b を，文字列 v と w にそれぞれ置き換えたときに，やはり文脈自由言語が得られるように思える．しかし，もしも出現するすべての a と b を，文脈自由な文字列集合 V と W にそれぞれ置き換えたとすると，それが文脈自由言語になるかどうかは，直感的には，さほど明らかではない．

注意点

今導入された概念の重要性を知るために，言語 $L = \{a^n b^n \mid n \in \mathbb{N}\}$ を考えてみよう．この言語は文脈自由であるから，直感的には $\{0^n 1^n \mid n \in \mathbb{N}\}$ もまた文脈自由である．なぜならば，我々は文法をたどりながら，出現するすべての a を 0 に，b を 1 に置き換えれば良いからである．

この観察から，もしも出現するすべての a と b を，文字列 v と w にそれぞれ置き換えたときに，やはり文脈自由言語が得られるように思える．しかし，もしも出現するすべての a と b を，文脈自由な文字列集合 V と W にそれぞれ置き換えたとすると，それが文脈自由言語になるかどうかは，直感的には，さほど明らかではない．
しかしながら, 我々はこの 閉包性を証明することを目指す． 以下では，代入の定義と同じように，常に 2 つの有限アルファベット Σ と Δ を仮定する．

代入に関する閉包 I

定理 1

$\mathcal{CF}$ は代入演算に関して閉じている．

代入に関する閉包 I

定理 1

$\mathcal{CF}$ は代入演算に関して閉じている．

(証明) ある 1 つの言語 $L \in \mathcal{CF}$ と ある代入 τ を考える．ただしすべての $a \in \Sigma$ に対して $\tau(a)$ は文脈自由言語であるとする．我々は $\tau(L)$ が文脈自由であることを示さなければならない．そのために，$L(\overline{\mathcal{G}}) = \tau(L)$ となるような文脈自由文法 $\overline{\mathcal{G}} = [\overline{T}, \overline{N}, \overline{\sigma}, \overline{P}]$ を作ろう．

代入に関する閉包 I

定理 1

$\mathcal{CF}$ は代入演算に関して閉じている．

(証明) ある 1 つの言語 $L \in \mathcal{CF}$ と ある代入 τ を考える．ただしすべての $a \in \Sigma$ に対して $\tau(a)$ は文脈自由言語であるとする．我々は $\tau(L)$ が文脈自由であることを示さなければならない．そのために，$L(\overline{\mathcal{G}}) = \tau(L)$ となるような文脈自由文法 $\overline{\mathcal{G}} = [\overline{T}, \overline{N}, \sigma, \overline{P}]$ を作ろう．

$L \in \mathcal{CF}$ であるから，$L = L(\mathcal{G})$ となるようなチョムスキー標準形の文脈自由文法 $\mathcal{G} = [\Sigma, N, \sigma, P]$ が存在する．次に，$\Sigma = \{a_1, \ldots, a_n\}$ として，すべての $a \in \Sigma$ に対して，$\tau(a)$ を考える．仮定より，すべての $a \in \Sigma$ に対して $\tau(a) \in \mathcal{CF}$ である．よって，すべての $a \in \Sigma$ に対して $\tau(a) = L(\mathcal{G}_a)$ となるような文脈自由文法 $\mathcal{G}_a = [T_a, N_a, \sigma_a, P_a]$ が存在する．ここで，集合 $N, N_{a_1}, \ldots, N_{a_n}$ が，互いに素であると仮定しても一般性は失われない．

代入に関する閉包 II

ここで，先に進むためのアイデアが必要である．それを得るため，$\mathcal{G}$ における可能な導出を調べてみよう．次の導出

$$\sigma \xRightarrow[\mathcal{G}]{*} x_1 x_2 \cdots x_m \ ,$$

をすべての $x_i \in \Sigma$ $(i = 1, \ldots, m)$ について考えよう．

代入に関する閉包 II

ここで，**先に進むためのアイデア**が必要である．それを得るため，$\mathcal{G}$ における可能な導出を調べてみよう．次の導出

$$\sigma \underset{\mathcal{G}}{\overset{*}{\Longrightarrow}} x_1 x_2 \cdots x_m \ ,$$

をすべての $x_i \in \Sigma$ $(i = 1, \ldots, m)$ について考えよう．すると，$\mathcal{G}$ はチョムスキー標準形であるから，生成規則 $(h_{x_i} \rightarrow x_i) \in P$, $i = 1, \ldots, m$, が必ず存在し，それによって以下の導出

$$\sigma \underset{\mathcal{G}}{\overset{*}{\Longrightarrow}} h_{x_1} h_{x_2} \cdots h_{x_m} \underset{\mathcal{G}}{\overset{m}{\Longrightarrow}} x_1 x_2 \cdots x_m \ , \tag{1}$$

が，すべての $h_{x_i} \in N$ について行われるはずである．

代入に関する閉包 III

像 (写像の行き先)$\tau(x_1 \cdots x_m)$ が

$$\tau(x_1)\tau(x_2)\cdots\tau(x_m)$$

によって計算できることを考慮すると，この像に含まれる各文字列 $w_1 w_2 \cdots w_m$ に対して，次の導出

$$\sigma_{x_i} \overset{*}{\underset{\mathcal{G}_{x_i}}{\Longrightarrow}} w_i \quad i = 1, \ldots, m\,.$$

が存在しなければならない．このことから直ちに $\overline{\mathcal{G}}$ を構築するアイデアが生まれる．

代入に関する閉包 IV

$h_{x_1}h_{x_2}\cdots h_{x_m}$ が得られたときに，式 (1) の 導出 を中断することを考えよう．$x_1x_2\cdots x_m$ を導出する代わりに必要なことは，$\sigma_{x_1}\cdots\sigma_{x_m}$ を生成することであり，すなわち，$i=1,\ldots,m$ について，生成規則 $(h_{x_i} \rightarrow x_i) \in P$ を $(h_{x_i} \rightarrow \sigma_{x_i}) \in \overline{P}$ で置き換える必要がある．したがって，以下を定義する．

代入に関する閉包 IV

$h_{x_1}h_{x_2}\cdots h_{x_m}$ が得られたときに，式 (1) の 導出 を中断することを考えよう．$x_1x_2\cdots x_m$ を導出する代わりに必要なことは，$\sigma_{x_1}\cdots\sigma_{x_m}$ を生成することであり，すなわち，$i=1,\ldots,m$ について，生成規則 $(h_{x_i} \to x_i) \in P$ を $(h_{x_i} \to \sigma_{x_i}) \in \overline{P}$ で置き換える必要がある．したがって，以下を定義する．

$$
\begin{aligned}
\overline{T} &= \bigcup_{a\in\Sigma} T_a \\
\overline{N} &= N \cup \left(\bigcup_{a\in\Sigma} N_a\right) \\
\overline{\sigma} &= \sigma \\
\overline{P} &= \left(\bigcup_{a\in\Sigma} P_a\right) \cup P[a /\!/ \sigma_a]\,.
\end{aligned}
$$

これを用いて $\overline{\mathcal{G}} = [\overline{T}, \overline{N}, \overline{\sigma}, \overline{P}]$ とする．

代入に関する閉包 V

$\tau(L) = L(\overline{\mathcal{G}})$ を示す作業が残っている.

主張 1. $\tau(L) \subseteq L(\overline{\mathcal{G}})$.

代入に関する閉包 V

$\tau(L) = L(\overline{\mathcal{G}})$ を示す作業が残っている.

主張 1. $\tau(L) \subseteq L(\overline{\mathcal{G}})$.

もし，$x_i \in \Sigma$ について $\sigma \overset{*}{\underset{\mathcal{G}}{\Longrightarrow}} x_1 \cdots x_m$ であり，$w_i \in T^*_{x_i}$ について $\sigma_{x_i} \overset{*}{\underset{\mathcal{G}_{x_i}}{\Longrightarrow}} w_i$ ならば（$i = 1 \ldots, m$），$x_1 \cdots x_m$ は

$$\sigma \overset{*}{\underset{\mathcal{G}}{\Longrightarrow}} h_{x_1} \cdots h_{x_m} \overset{*}{\underset{\mathcal{G}}{\Longrightarrow}} x_1 \cdots x_m \,,$$

と導出される（ただし $h_{x_i} \in N$）．すると明らかに,

$$\sigma \overset{*}{\underset{\overline{\mathcal{G}}}{\Longrightarrow}} h_{x_1} \cdots h_{x_m} \overset{*}{\underset{\overline{\mathcal{G}}}{\Longrightarrow}} \sigma_{x_1} \cdots \sigma_{x_m} \overset{*}{\underset{\overline{\mathcal{G}}}{\Longrightarrow}} w_1 \cdots w_m \,.$$

を生成できる．以上により主張 1 は示された.

代入に関する閉包 VI

主張 2. $L(\overline{\mathcal{G}}) \subseteq \tau(L)$.

今度は $\sigma \overset{*}{\Rightarrow} w$ $(w \in \overline{T}^*)$ から始めよう．ただし，$w \in \overline{T}^*$ である．もし $w = \lambda$ ならば，$\sigma \rightarrow \lambda$ も P に含まれるので主張は示される．

代入に関する閉包 VI

主張 2. $L(\overline{\mathcal{G}}) \subseteq \tau(L)$.

今度は $\sigma \stackrel{*}{\Rightarrow} w$ ($w \in \mathsf{T}^*$) から始めよう．ただし，$w \in \mathsf{T}^*$ である．もし $w = \lambda$ ならば，$\sigma \to \lambda$ も P に含まれるので主張は示される．そうでなければ，$\overline{\mathcal{G}}$ を構築すれば，w の導出が以下の通りであることが確認される．

$$\sigma \underset{\overline{\mathcal{G}}}{\overset{*}{\Longrightarrow}} \sigma_{x_1} \cdots \sigma_{x_m} \underset{\overline{\mathcal{G}}}{\overset{*}{\Longrightarrow}} w\,.$$

代入に関する閉包 VI

主張 2. $L(\overline{\mathcal{G}}) \subseteq \tau(L)$.

今度は $\sigma \overset{*}{\Rightarrow} w$ ($w \in \mathsf{T}^*$) から始めよう．ただし，$w \in \mathsf{T}^*$ である．もし $w = \lambda$ ならば，$\sigma \to \lambda$ も P に含まれるので主張は示される．そうでなければ，$\overline{\mathcal{G}}$ を構築すれば，w の導出が以下の通りであることが確認される．

$$\sigma \underset{\overline{\mathcal{G}}}{\overset{*}{\Longrightarrow}} \sigma_{x_1} \cdots \sigma_{x_m} \underset{\overline{\mathcal{G}}}{\overset{*}{\Longrightarrow}} w\,.$$

構成法より，式 (1) で示した通り，$\sigma \underset{\mathcal{G}}{\overset{*}{\Longrightarrow}} x_1 \cdots x_m$ であることがわかっている．
さらに，すべての $i = 1, \ldots, m$ に対して，$w = w_1 \cdots w_m$ かつ $\sigma_{x_i} \underset{\mathcal{G}_{x_i}}{\overset{*}{\Longrightarrow}} w_i$ となるような文字列 $w_1, \ldots, w_m \in \mathsf{T}^*$ が存在する．

代入に関する閉包 VII

以上より $w_i \in \tau(x_i)$ であり，よって $w \in \tau(L)$ が示された．

代入に関する閉包 VII

以上より $w_i \in \tau(x_i)$ であり，よって $w \in \tau(L)$ が示された．

最後に，主張 1 と 2 を合わせて，

$$\tau(L) = L(\overline{\mathcal{G}})$$

を得る． ∎

代入に関する閉包 VII

以上より $w_i \in \tau(x_i)$ であり，よって $w \in \tau(L)$ が示された．

最後に，主張 1 と 2 を合わせて，

$$\tau(L) = L(\overline{\mathcal{G}})$$

を得る． ∎

我々の定理は，次の好ましい系を持つ．

系 2

$\mathcal{CF}$ は準同型写像に関して閉じている．

(証明) 準同型写像は代入の特殊な形なので，すべての単一要素の部分集合が文脈自由であることを議論すれば十分である．しかしこれは自明である．なぜならば，すべての有限な言語は $\mathcal{REG}$ に属することと，$\mathcal{REG} \subseteq \mathcal{CF}$ であることは，すでに証明している．よってこの系は正しい． ∎

ダイク言語 I

文脈自由言語の講義を始める際に強調したことは，多くのプログラミング言語では，複数の種類のバランスの取れた括弧を使用するということである．したがって，括弧からなる言語をもっと詳しく見ていこう．そのような言語はダイク言語 (*Dyck languages*) と呼ばれている．

ダイク言語 I

文脈自由言語の講義を始める際に強調したことは，多くのプログラミング言語では，複数の種類のバランスの取れた括弧を使用するということである．したがって，括弧からなる言語をもっと詳しく見ていこう．そのような言語はダイク言語 (*Dyck languages*) と呼ばれている．

ダイク言語を定義するために，以下の記法が必要である．
$n \in \mathbb{N}^+$ とし，

$$X_n = \{a_1, \overline{a}_1, a_2, \overline{a}_2, \ldots, a_n, \overline{a}_n\}.$$

とする．X_n は，相異なる括弧記号の集合であって，a_i は開く括弧，$\overline{a}_i$ はそれに対応する閉じる括弧である．つまり X_n は n 種類の異なる括弧記号の集合である．

ダイク言語 II

以上でダイク言語を定義する準備が整った.

定義 4

次の条件を満たすとき，言語 L は n 種類の括弧記号からなる**ダイク言語 (Dyck language)** であると言う．その条件は，文法 $\mathcal{G}_n = [X_n, \{\sigma\}, \sigma, P_n]$，ただし，

$$P_n = \{\sigma \rightarrow \lambda,\ \sigma \rightarrow \sigma\sigma,\ \sigma \rightarrow a_1\sigma\overline{a}_1,\ \ldots,\ \sigma \rightarrow a_n\sigma\overline{a}_n\}$$

により生成される言語 D_n と，L が同型 *(isomorphic)* なことである.

ダイク言語 II

以上でダイク言語を定義する準備が整った．

定義 4

次の条件を満たすとき，言語 L は n 種類の括弧記号からなる**ダイク言語 (Dyck language)** であると言う．その条件は，文法 $\mathcal{G}_n = [X_n, \{\sigma\}, \sigma, P_n]$，ただし，

$$P_n = \{\sigma \to \lambda,\ \sigma \to \sigma\sigma,\ \sigma \to a_1\sigma\bar{a}_1,\ \ldots,\ \sigma \to a_n\sigma\bar{a}_n\}$$

により生成される言語 D_n と，L が同型 *(isomorphic)* なことである．

ダイク言語の重要性は，この後すぐ明らかになる．なぜならば，我々はこれを用いて，文脈自由言語の美しい特徴を示す定理を証明するからである．

チョムスキー・シュツェンベルガーの定理 I

定理 3 (Chomsky-Schützenberger の定理)

すべての文脈自由言語 L に対して，以下の式が成り立つような，$n \in \mathbb{N}^{+}$ と準同型写像 h と正規言語 R_L が存在する．

$$L = h(D_n \cap R_L) .$$

チョムスキー・シュツェンベルガーの定理 II

(証明) 任意の 1 つの文脈自由言語 L を考える．ここで $\lambda \notin L$ を仮定しても一般性は失われない．さらに，$L = L(\mathcal{G})$ であるようなチョムスキー標準形の文脈自由文法 $\mathcal{G} = [T, N, \sigma, P]$ を考える．$T = \{x_1, \ldots, x_m\}$ として，P に含まれるすべての生成規則を考える．$\mathcal{G}$ はチョムスキー標準形であるから，すべての生成規則は $h_i \to h_i' h_i''$ または $h_j \to x$ の形をしている．非終端の生成規則（すなわち $h_i \to h_i' h_i''$）の総数を t とする．なお，そのような生成規則から任意の 2 つを選んだとき，全部ではないがいくつかの非終端記号が重複することは，当然あり得る．

チョムスキー・シュツェンベルガーの定理 III

全部で m 個の終端記号と t 個の生成規則がある．そこで，以下のアルファベットに対するダイク言語 D_{m+t} を作ってみよう．

$$\begin{aligned} X_{m+t} &= \{\overline{x}_1, \ldots, \overline{x}_m, \overline{x}_{m+1}, \ldots, \overline{x}_{m+t}, \\ &\quad x_{m+1}, \ldots, x_{m+t}, x_1, \ldots, x_m\}\,. \end{aligned}$$

チョムスキー・シュツェンベルガーの定理 III

全部で m 個の終端記号と t 個の生成規則がある．そこで，以下のアルファベットに対するダイク言語 D_{m+t} を作ってみよう．

$$\begin{aligned} X_{m+t} &= \{\overline{x}_1, \ldots, \overline{x}_m, \overline{x}_{m+1}, \ldots, \overline{x}_{m+t}, \\ &\quad x_{m+1}, \ldots, x_{m+t}, x_1, \ldots, x_m\}. \end{aligned}$$

次に，以下の通り定義される写像 $\chi_{m+t} \colon X_{m+t} \longrightarrow T^*$ を考える．

$$\chi_{m+t}(x_j) = \begin{cases} x_j, & \text{ただし } 1 \leqslant j \leqslant m\,; \\ \lambda, & \text{ただし } m+1 \leqslant j \leqslant m+t\,. \end{cases}$$

そして，すべての $j = 1, \ldots, m+t$ に対し $\chi_{m+t}(\overline{x}_j) = \lambda$ とする．
χ_{m+t} が準同型写像であることの証明は，演習問題として残しておく．

チョムスキー・シュツェンベルガーの定理 IV

ここまでで，以下の文法を定義する準備ができた．
$\mathcal{G}_L = [X_{m+t}, N, \sigma, P_L]$, ただし

$$
\begin{aligned}
P_L &= \{h \to x_i \bar{x}_i \mid 1 \leqslant i \leqslant m \text{ かつ } (h \to x_i) \in P\} \\
&\cup \{h \to x_i \bar{x}_i \bar{x}_{m+j} h_j'' \mid 1 \leqslant i \leqslant m,\ (h \to x_i) \in P,\ 1 \leqslant j \leqslant t\} \\
&\cup \{h_j \to x_{m+j} h_j' \mid 1 \leqslant j \leqslant t\}\,.
\end{aligned}
$$

明らかに $\mathcal{G}_L$ は 正規文法である．そこで $R_L = L(\mathcal{G}_L)$ とおいて，

$$L = \chi_{m+t}(D_{m+t} \cap R_L)\,.$$

を証明することを目指そう．これはこの後の主張と補題により示される．

チョムスキー・シュツェンベルガーの定理 V

主張 1. $L \subseteq \chi_{m+t}(D_{m+t} \cap R_L)$.
主張 1 の証明は主に以下の補題によって示される.

補題 4

$\mathcal{G}$ は上記の L を生成する文法であるとし，$\mathcal{G}_L$ は R_L を生成する文法であるとする．そして $h \in N$ とする．もしも

$$h \underset{\mathcal{G}}{\overset{1}{\Longrightarrow}} w_1 \underset{\mathcal{G}}{\overset{1}{\Longrightarrow}} w_2 \underset{\mathcal{G}}{\overset{1}{\Longrightarrow}} \cdots \underset{\mathcal{G}}{\overset{1}{\Longrightarrow}} w_{n-1} \underset{\mathcal{G}}{\overset{1}{\Longrightarrow}} w_n \in T^*$$

ならば，$h \underset{\mathcal{G}_L}{\overset{*}{\Longrightarrow}} q$ かつ $\chi_{m+t}(q) = w_n$ となるような $q \in D_{m+t}$ が存在する.

チョムスキー・シュツェンベルガーの定理 VI

この補題は，導出の長さ n に関する数学的帰納法で証明される．
帰納法の基底として $n = 1$ とおく．すると，

$$h \overset{1}{\underset{\mathcal{G}}{\Longrightarrow}} w_1 \in T^*$$

と仮定される．$\mathcal{G}$ はチョムスキー標準形であるから，
$(h \ \to \ w_1) \in P$ が言える．したがって，チョムスキー標準形の定義より，ある $x \in T$ に対して $w_1 = x$ でなければならない．

チョムスキー・シュツェンベルガーの定理 VII

我々は，$h \overset{*}{\underset{\mathcal{G}_L}{\Longrightarrow}} q$ かつ $\chi_{m+t}(q) = x$ であるような $q \in D_{m+t}$ が存在することを示さなければならない．定義より明らかに，生成規則 $h \rightarrow x\overline{x}$ は P_L に属している（P_L の定義の第 1 の集合を参照）．よって，単に $q = x\overline{x}$ としてよい．すると帰納法の基底が示される．なぜならば χ_{m+t} の定義から直ちに以下が導かれるからである．

$$\chi_{m+t}(q) = \chi_{m+t}(x\overline{x}) = \chi_{m+t}(x)\chi_{m+t}(\overline{x}) = x\lambda = x\,.$$

チョムスキー・シュツェンベルガーの定理 VIII

$n \geqslant 1$ における帰納法の仮定に基づき，$n+1$ に対する帰納ステップを進めよう．そこで，

$$h \underset{\mathcal{G}}{\overset{1}{\Longrightarrow}} w_1 \underset{\mathcal{G}}{\overset{1}{\Longrightarrow}} \cdots \underset{\mathcal{G}}{\overset{1}{\Longrightarrow}} w_n \underset{\mathcal{G}}{\overset{1}{\Longrightarrow}} w_{n+1} \in T^*$$

という長さ $n+1$ の導出を考える．

チョムスキー・シュツェンベルガーの定理 VIII

$n \geqslant 1$ における帰納法の仮定に基づき，$n+1$ に対する帰納ステップを進めよう．そこで，

$$h \xrightarrow[\mathcal{G}]{1} w_1 \xrightarrow[\mathcal{G}]{1} \cdots \xrightarrow[\mathcal{G}]{1} w_n \xrightarrow[\mathcal{G}]{1} w_{n+1} \in \mathsf{T}^*$$

という長さ $n+1$ の導出を考える．
$n \geqslant 1$ であって，導出の長さが少なくとも 2 であるから，w_1 を導出するための生成規則は $h \to h'h''$（ただし $h, h', h'' \in \mathsf{N}$）という形でなければならない．したがって，ある j が存在して，$1 \leqslant j \leqslant t$ かつ $h = h_j$ かつ $w_1 = h_j'h_j''$ とならなければならない．

チョムスキー・シュツェンベルガーの定理 IX

以上の考察より，ある v_1, v_2 が存在して，$w_{n+1} = v_1 v_2$ かつ

$$h'_j \underset{\mathcal{G}}{\overset{*}{\Longrightarrow}} v_1 \quad \text{かつ} \quad h''_j \underset{\mathcal{G}}{\overset{*}{\Longrightarrow}} v_2$$

となるはずである．完全な導出の長さが $n+1$ であるから，v_1 と v_2 の導出の長さは，どちらも n 以下でなければならない．したがって，帰納法の仮定を適用できる．すなわち，$q_1, q_2 \in D_{m+t}$ かつ $\chi_{m+t}(q_1) = v_1$ かつ $\chi_{m+t}(q_2) = v_2$ であるような文字列 q_1 と q_2 が存在する．

チョムスキー・シュツェンベルガーの定理 X

さらに，帰納法の仮定から，以下がわかる．

$$h'_j \overset{*}{\underset{\mathcal{G}_L}{\Longrightarrow}} q_1 \quad \text{かつ} \quad h''_j \overset{*}{\underset{\mathcal{G}_L}{\Longrightarrow}} q_2 \,.$$

$(h_j \rightarrow h'_j h''_j) \in P$ であることを考慮すると，明らかに，生成規則 $h_j \rightarrow x_{m+j} h'_j$ は P_L に含まれる．よって，

$$h = h_j \overset{1}{\underset{\mathcal{G}_L}{\Longrightarrow}} x_{m+j} h'_j \overset{*}{\underset{\mathcal{G}_L}{\Longrightarrow}} x_{m+j} q_1$$

は正規文法の導出である．さらに，この導出の最後のステップは，次の形をしているはずである．

$$x_{m+j} q'_1 h_k \overset{1}{\underset{\mathcal{G}_L}{\Longrightarrow}} x_{m+j} q'_1 x\overline{x} \,.$$

ただし生成規則 $h_k \rightarrow x\overline{x}$ が適用されており，x は条件 $q_1 = q'_1 x\overline{x}$ を満たすものとする．

チョムスキー・シュツェンベルガーの定理 XI

ここで最後のステップを，これもまた P_L に属する 生成規則 $h_k \rightarrow x\overline{x}x_{m+j}h_j''$ で置き換えると，

$$h = h_j \underset{\mathcal{G}_L}{\overset{1}{\Longrightarrow}} x_{m+j}h_j' \underset{\mathcal{G}_L}{\overset{*}{\Longrightarrow}} x_{m+j}q_1\overline{x}_{m+j}h_j''$$

$$\underset{\mathcal{G}_L}{\overset{*}{\Longrightarrow}} x_{m+j}q_1\overline{x}_{m+j}q_2 =: q \in D_{m+t} .$$

が得られる．D_{m+t} に属するのは，ダイク言語の定義に加え，q_1 の回りに括弧 x_{m+j} と $\overline{x}_{m+j}$ が正しく使われていること，および，$q_2 \in D_{m+t}$ であることによる．最後に χ_{m+t} の定義より $\chi_{m+t}(x_{m+j}q_1\overline{x}_{m+j}q_2) = v_1v_2$ であることは確かである．以上により補題が証明され，$h = \sigma$ に対して直ちに主張 1 が示される．

チョムスキー・シュツェンベルガーの定理 XII

主張 2. $L \supseteq \chi_{m+t}(D_{m+t} \cap R_L)$.

これもまた，次に述べる補題により証明する．

補題 5

$\mathcal{G}$ は先に定めた L を生成する文法であるとし，$\mathcal{G}_L$ は R_L を生成する文法であるとする．そして $h \in N$ とする．もしも

$$h \underset{\mathcal{G}_L}{\overset{1}{\Longrightarrow}} w_1 \underset{\mathcal{G}_L}{\overset{1}{\Longrightarrow}} \cdots \underset{\mathcal{G}_L}{\overset{1}{\Longrightarrow}} w_n \in D_{m+t}$$

ならば $h \underset{\mathcal{G}}{\overset{*}{\Longrightarrow}} \chi_{m+t}(w_n)$ である．

チョムスキー・シュツェンベルガーの定理 XIII

この補題は，導出の長さに関する数学的帰納法で証明される．帰納法の基底として $n = 1$ とする．

$$h \overset{1}{\underset{\mathcal{G}_L}{\Longrightarrow}} w_1 \in D_{m+t} \ .$$

を考えると，$(h \to w_1) \in P_L$ となるはずである．したがって，$w_1 = x_i\overline{x}_i, 1 \leqslant i \leqslant m$ かつ $(h \to x_i\overline{x}_i) \in P_L$ となるような $x_i\overline{x}_i$ が必ず存在する．P_L の定義より，$(h \to x_i) \in P$ が言える．したがって，

$$h \overset{1}{\underset{\mathcal{G}}{\Longrightarrow}} x_i = \chi_{m+t}(x_i\overline{x}_i) = \chi_{m+t}(w_1) \ .$$

以上で帰納法の基底が示された．

チョムスキー・シュツェンベルガーの定理 XIV

帰納ステップはテキストに示されている．

主張 2 は，$h = \sigma$ に対する補題より直接導かれる．

主張 1 と主張 2 を合わせて，定理が示された．

Thank you!