情　報　学　基　礎 38－1
（１９９５．　７．　２８）

確率文法におけるパラメタの一致推定

日高 達

E-mail: hitaka@lang.ai.kyushu-u.ac.jp

九州大学 工学部 情報工学科

〒812-81　福岡市東区箱崎 6-10-1

確率文法のパラメタは，収集された標本集合を用いて推定されるが，この推定は一致推定であることが望ましい．一致推定とは，概略的に言うと,「標本数が大きいと推定値は真値に十分近いことが期待される」ことを意味している．本論文では，(1) 一致推定を得ることが困難な，病的な確率文脈自由文法のクラスが存在することを示し，(2) 病的な確率文脈自由文法の特徴づけを行ない，(3) 健全な確率文脈自由文法の制限されたクラスに対するパラメタの一致推定式を与える．

Consistent Estimate for Parameters of Probabilistic Grammars

Toru Hitaka

Department of Computer Science and Communication Engineering
Faculty of Engineering, Kyushu University

10-1, 6-chome, Hakozaki, Higasi-ku, Fukuoka, 812-81 Japan

Values of parameters in a probabilistic grammar are estimated from a set of samples. Parameter estimate is desirable to be consisitent which roughly means that if the number of samples is large enough, esimated each value is very close to true value. It is shown that there is a morbid class of Probabilistic Context-Free Grammars whose consisitent estimate is difficult to get. Then, characterization of this morbid class and consistent estimate for a restricted class of sound PCFGs are given.

1　はじめに

最近の計算機環境の向上と大量の事例データの蓄積により，言語処理に事例データを反映する方法が模索されている．事例データを言語処理に反映する方法は，古くは確率有限オートマトンとしてモデル化され，文字認識や音声認識に応用されており，最近では，確率有限オートマトンの自然な拡張として，確率文脈自由文法（Probabilistic Context Free Grammar）がテキスト認識や音声認識における数学モデルとして広く用いられている．これらの確率モデルは数多くの確率パラメタを含んでおり，これらのパラメタは事例データ（標本）の統計的性質に基づいて推定される．概略的には標本集合が最も発生しやすいような値をパラメタ値として推定するのである．

確率パラメタの推定には，統計学の立場から，望ましいとされる幾つかの基準がある．なかでもとりわけ重要な要請が一致性（Consistency）である．一致推定とは，標本の数を無限に大きくしていくと，推定値が真値に（確率 1 で）収束（確率収束）するということである．

確率有限オートマトンにおける状態遷移確率や確率文脈自由文法おける書換え規則の適用確率などの，いわゆる確率パラメタは，従来，標本集合に関する関数，例えばその標本集合が発生する確率など，を目的関数として導入し，目的関数の値を最大にするようなパラメタ値を推定値とすることで満足し，一致推定を求める研究は私の知る限りなされていない．

本論文は，確率文脈自由文法の確率パラメタの一致推定法を求めることを目的とする．本論文の成果は次の 4 点に集約される．

1 すべての構文木（または，すべての文）の生起確率の総和を求める式を与えた．

2 構文木 T において，書換え規則 δ が適用された回数を $c(\delta,T)$ とすると，$c(\delta,T)$ の一次モーメント ($c(\delta,T)$ の期待値) と二次モーメント ($c^2(\delta,T)$ の期待値) を求める式を与えた．これらの式より，$c(\delta,T)$ の一次モーメントが存在することと二次モーメントが存在することは同値であることが示される．したがって，$c(\delta,T)$ の期待値が存在すれば，その分散も存在することになる．

3 左辺を同一の非終端記号とする書換え規則を δ,δ' とする．すべての構文木の生起確率の総和が 1 である確率文脈自由文法（健全な確率文脈自由文法と呼ぶ）の場合には，$c(\delta,T)$ の期待値と $c(\delta',T)$ の期待値の比は δ の適用確率と δ' の適用確率の比に等しいことが示される．

4 健全な確率文脈自由文法において，すべての書換え規則 δ について，$c(\delta,T)$ の期待値が存在する場合に，確率パラメタの一致推定を与えた．

$c(\delta,T)$ の期待値が無限になる場合を含めた，健全な確率文脈自由文法の一致推定に関しては既に見通しを得ており，近く発表する．

2　確率文脈自由文法

確率文脈自由文法 PCFG(Probabilistic Context Free Grammar) の定義を与える．ただし，文脈自由文法についての知識は持っているものとする．

【定義 1】 本論文では確率文脈自由文法 G を，次のような 5 組で定義する．

$$G = (\Sigma, V, P, S, p)$$

$\sum$: 終端記号 (a,b 等の小文字で表す) の有限集合.
V : 非終端記号 $X_1, X_2, \cdots, X_n$ の集合.
X_1 : 開始記号 ($X_1 \in V$).
P : 書換え規則の有限集合.
$p : P \to (0,1]$　(P から実数の区間 $(0,1]$ への写像).

本論文では，書換え規則を (1) のように表す．また，適用確率には (2) が成立する．

$$X_i \to \alpha_{ij} \quad (i = 1,2,\cdots,n,\ \ j = 1,2,\cdots,m_i) \tag{1}$$

$$p(X_i \to \alpha_{i1}) + p(X_i \to \alpha_{i2}) + \cdots + p(X_i \to \alpha_{im_i}) = 1. \quad (i = 1,2,\cdots,n,\ \ j = 1,2,\cdots,m_i) \tag{2}$$

(1) の意味はおよそ次のようである．すなわち，文の生成過程を文脈自由文法による確率的生成過程と見る．生成途中の文形式に出現する非終端記号 X_i は，次に X_i を左辺とする書換え規則の適用を受けるわけであるが，その場合，$X_i \to \alpha_{ij}$ の適用を受けて X_i が α_{ij} に書換わる確率が $p(X_i \to \alpha_{ij})$ である．また，$X_i, \alpha_{ij}, p(X_i \to \alpha_{ij})$ を上の書換え規則のそれぞれ左辺，右辺，適用確率と言う．以後，$X_i \to \alpha_{ij}$ を δ_{ij} で表す．

確率以外の文，言語 $L(G)$, 導出（木），文形式，最左（右）導出等の定義は CFG の場合と全く同じなので省略する．

SCFG では文 $s \in L(G)$ と s の導出木 Tに対し，それぞれ s の生起確率 $P_r(s)$ と T の生起確率 $P_r(T)$ が次のように定義される．すなわち，$P_r(s)$ は文 s を導出するすべての導出木の生起確率の総和であり，$P_r(T)$ は T の導出に適用された書換え規則の適用確率の（重複を含めた）積である．

さらに，本質的でない議論を避けるために，本論文で扱う PCFG は，無駄な非終端記号を含まない(well formed), すなわち，任意の $X_i \in V$ に対して，$\omega \in \Sigma^*$ と $\alpha \in (\Sigma \cup V)^*, \beta \in (\Sigma \cup V)^*$が存在し，

$$X_i \overset{+}{\Longrightarrow} \omega, \quad X_1 \overset{*}{\Longrightarrow} \alpha X_i \beta. \tag{3}$$

が成立するものとする． □

3 確率パラメタの推定

PCFG $G = (\Sigma, V, P, X_1, p)$ において，$p : P \to (0, 1]$ は標本(事例データ)集合に依存して定められるパラメタである．

【定義 2】 標本集合
ランダムに収集された標本(構文木)を $T_1, T_2, \cdots, T_N$ とし，$\delta \in P$が構文木 T の導出に適用された回数を $c(\delta, T)$ で表す．ここで，N 個の標本の列 $\mathcal{F}_N = (T_1, T_2, \cdots, T_N)$ が収集される確率 $P_r(\mathcal{F}_N)$ は次のようになる．

$$P_r(\mathcal{F}_N) = \prod_{k=1}^{N} P_r(T_k) \tag{4}$$

$\mathcal{F}_N = (T_1, T_2, \cdots, T_N)$ に基づいて，パラメタ $p(\delta)$ $(\delta \in P)$ の推定がなされるが，この推定値を$\widehat{p}(\delta|\mathcal{F}_N)$ と記す． □

パラメタの推定値は，N が大きければ $\widehat{p}(\delta|\mathcal{F}_N)$ は真値 $p(\delta)$ に確率収束することが望ましい．

【定義 3】 一致推定
任意の $\delta \in P$ と任意の $\varepsilon (> 0)$ に対し，

$$\lim_{N \to \infty} P_r\{|\widehat{p}(\delta|\mathcal{F}_N) - p(\delta)| < \varepsilon\} = 1 \tag{5}$$

であるとき，そのパラメタ推定は一致推定 (consistent estimate) であると言う．ただし，$P_r\{|\widehat{p}(\delta|\mathcal{F}_N) - p(\delta)| < \varepsilon\}$ は，推定値$\widehat{p}(\delta|\mathcal{F}_N)$ と真値 $p(\delta)$ の差が $\pm\varepsilon$ 以内であるような標本列 $\mathcal{F}_N$ が収集される確率である． □

一致性が意味することは，標本の個数が大きければ，推定値が真値に近くなることが確率的に保証されるということであり，パラメタ推定上極めて重要な要請と考えられる．

標本 $T_1, \cdots, T_N$ が収集される(または発生する)確率(尤度)を最大にする方法を一般に最尤推定 (maximum likelihood estimate) と言う．

4 文法の健全性と特性方程式

PCFG において，すべての構文木の生起確率の総和は，1 になるとは限らない．1 になる PCFG は**健全** (sound) であるといい，そうでない PCFG は**不健全** (morbid) であるという．

5 章で不健全な PCFG に対しては，確率パラメタの一致推定が困難であることが示される．

T.Booth と J.Thompson は文献 [1] で，任意に与えられた PCFG が健全であるための十分条件を次のように与えている．

【X_i の子に X_j が現れる個数の期待値を (i, j) 要素とする $n \times n$ 行列 A(expectation matrix) のすべての固有値の絶対値が 1 より小さい.】

例 1. $G = (\Sigma, V, P, X_1, p)$

$$\Sigma = \{a, b\}, \quad V = \{X_1, X_2\},$$

$$P = \{X_1 \xrightarrow[\frac{1}{3}]{} aX_1X_2bX_2, X_1 \xrightarrow[\frac{2}{3}]{} ab, X_2 \xrightarrow[1]{} bX_1\}.$$

例 1 の PCFG では，expectation matrix A は

$$A = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} \\ 1 & 0 \end{bmatrix}.$$

A の固有値は $-\frac{2}{3}, 1$ となり，Booth & Thompson の条件を満足しない．この PCFG が健全であることが，後で示される．

【定義 4】 X_i を root とする構文木を X_i からの構文木と呼び，開始記号 X_1 からの構文木を簡略に構文木と呼ぶ．X_i からの構文木の中で，root X_i から leaf までの path length がすべて d 以下（以後，深さ d 以下と呼ぶ）の構文木の全体集合を $T_d(X_i)$ と記し，最初の書換え規則に $X_i \to \alpha_{ij} \in P$ を適用した深さ d 以下の X_i からの構文木の全体集合を $T_d(X_i \to \alpha_{ij})$ と記す．$T_{d+1}(X_i \to \alpha_{ij})$

に属する構文木は，root X_i の子である α_{ij} の各非終端記号 X_k の節に $\mathcal{T}_d(X_k)$ に属する構文木を連結した木である．また，X_i からの構文木の全体集合を $\mathcal{T}(X_i)$ と記し，最初の書換え規則に $X_i \to \alpha_{ij} \in P$ を適用した X_iからの構文木の全体集合を $\mathcal{T}(X_i \to \alpha_{ij})$ と記す．また，$\mathcal{T}_d(X_i)$ に属するすべての木の生起確率 $P_r(\mathcal{T}_d(X_i))$ を i 次要素とする n 次元ベクトルを W_dとし，$d \to \infty$ での極限値を W と記す．

$$W_d = \begin{bmatrix} P_r(\mathcal{T}_d(X_1)) \\ \vdots \\ P_r(\mathcal{T}_d(X_n)) \end{bmatrix}, \quad W_\infty = \lim_{d\to\infty} W.$$

□

$$\left.\begin{aligned} &\mathcal{T}_d(X_i) \subseteq \mathcal{T}_{d+1}(X_i) \\ &\mathcal{T}_d(X_i \to \alpha_{ij}) \subseteq \mathcal{T}_{d+1}(X_i \to \alpha_{ij}) \end{aligned}\right\} \qquad (6)$$

$$\left.\begin{aligned} &\mathcal{T}(X_i) = \lim_{d\to\infty} \mathcal{T}_d(X_i) \\ &\mathcal{T}(X_i \to \alpha_{ij}) = \lim_{d\to\infty} \mathcal{T}_d(X_i \to \alpha_{ij}) \end{aligned}\right\} \qquad (7)$$

$$\left.\begin{aligned} &\mathcal{T}_d(X_i) = \bigcup_{j=1}^{m_i} \mathcal{T}_d(X_i \to \alpha_{ij}) \\ &\mathcal{T}(X_i) = \bigcup_{j=1}^{m_i} \mathcal{T}(X_i \to \alpha_{ij}) \end{aligned}\right\} \qquad (8)$$

【定義 5】 特性方程式　$X = F_G(X)$

$\alpha \in (\Sigma \cup V)^*$ において，α に現れるすべての終端記号を整数 1 で置換したものを $\overline{\alpha}$ で 表す．$\overline{\alpha}$ における各 X_iを実数の上を動く変数と解釈すると，$\overline{\alpha}$ は実数変数の積と解釈される．$PCFG\ G = (\Sigma, V, P, X_1, p)$ に対し，$X_1, X_2, \cdots, X_n$ の多項式 $f_i(X), X, F_G(X)$ を次のように定義する．

$$f_i(X) = p(\delta_{i1}) \cdot \overline{\alpha}_{i1} + p(\delta_{i2}) \cdot \overline{\alpha}_{i2} + \cdots + p(\delta_{im_i}) \cdot \overline{\alpha}_{im_i} \quad (i = 1, 2, \cdots, n)$$

$f_i(X)$ は係数の総和が 1 であるような，正係数の多項式になっている．

$$X = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}, \quad F_G(X) = \begin{bmatrix} f_1(X) \\ \vdots \\ f_n(X) \end{bmatrix}.$$

このとき，次の (9) 式を G の 特性方程式と呼ぶ．

$$X = F_G(X) \qquad (9)$$

特性方程式 (9) は，次の連立方程式の簡略表記である．

$$X_i = f_i(X) \quad (i = 1, 2, \cdots, n)$$

$F_G(X)$ の i 行目の多項式 $(= f_i(X))$ を $[F_G(X)]_i$ と記す．非負整数 k に対し，$F_G^k(X)$ を次のように帰納的に定義する．

$$\left.\begin{aligned} &F_G^0(X) \stackrel{\text{def}}{=} X \\ &F_G^{k+1}(X) \stackrel{\text{def}}{=} F_G(F_G^k(X)) \quad (k = 0, 1, 2, \cdots) \end{aligned}\right\} \qquad (10)$$

また，誤解の恐れのない限り，F_Gを簡略に Fと記す　□

例 1 の PCFG にたいしては，

$$F(X) = \begin{bmatrix} \frac{1}{3}X_1X_2^2 + \frac{2}{3} \\ X_1 \end{bmatrix},$$

$$\begin{aligned} F^2(X) &= \begin{bmatrix} \frac{1}{3}(\frac{1}{3}X_1X_2^2 + \frac{2}{3})X_1^2 + \frac{2}{3} \\ (\frac{1}{3}X_1X_2^2 + \frac{2}{3}) \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{9}X_1^3X_2^2 + \frac{2}{9}X_1^2 + \frac{2}{3} \\ \frac{1}{3}X_1X_2^2 + \frac{2}{3} \end{bmatrix} \end{aligned}$$

$[F^k(X)]_i$ は，係数の総和が 1 になるような，正値係数の多項式であることに注意しよう．

n 次元ベクトル A, B において，

$$[A]_i \leq [B]_i \quad (i = 1, 2, \cdots n)$$

であることを，$A \leq B$ と記す．$F_G(X)$ に対し，次の補題が成立する．

【補題 4.1】 $F(X)$ に対し，次の i〜iii が成立する．

i $0 \leq A \leq B$に対し，$F(A) \leq F(B)$.（単調性）

ii $A = F(A)$, かつ $A \geq 0$ ならば，任意の非負整数 kに対し，$F^k(0) \leq A$.

iii 任意の非負整数 kに対し，$0 \leq F^k(0) \leq F^{k+1}(0) \leq \mathbf{1}$. ここで，0 は 0 ベクトル，$\mathbf{1}$ は各要素が 1 のベクトルである．

証明の概略．i は，各 $[F_G(X)]_i$ が正係数の多項式であることから自明．ii は i を用いて，k に関する数学的帰納法により証明される．iii は，$F_G(\mathbf{1}) = \mathbf{1}$ であることと，i を用いて，k に関する数学的帰納法により証明される．□

$\alpha \in (\Sigma \cup V)^*$ に現れる非終端記号で，左方から数えて k 番目の非終端記号を $X_{k(\alpha)}$ と記す．つまり，α における左から k番目の非終端記号の index を $k(\alpha)$ で表す．例えば，$\alpha = aX_1X_2bX_2$の場合，$1(\alpha) = 1, 2(\alpha) = 2, 3(\alpha) = 2$ である．

【 補題 4.2 】

i $W_{d+1} = F(W_d)$

ii $W_d = F^d(0)$

証明. $\mathcal{T}_{d+1}(X_i \to \alpha_{ij})$ に属する木は, root X_i の子 α_{ij} における各非終端ノード $X_{k(\alpha_{ij})}$ に $\mathcal{T}_d(X_{k(\alpha_{ij})})$ に属する木を接続したものである. したがって,

$$
\begin{aligned}
& P_r(\mathcal{T}_{d+1}(X_i \to \alpha_{ij})) \\
&= p(\delta_{ij}) \sum_{\substack{T_1 \in \mathcal{T}_d(X_{1(\alpha_{ij})}) \\ T_2 \in \mathcal{T}_d(X_{2(\alpha_{ij})})}} P_r(T_1) \cdot P_r(T_2) \cdot \cdots \\
& \quad \vdots \\
&= p(\delta_{ij}) \cdot \sum_{T_1 \in \mathcal{T}_d(X_{1(\alpha_{ij})})} P_r(T_1) \\
& \qquad \cdot \sum_{T_2 \in \mathcal{T}_d(X_{2(\alpha_{ij})})} P_r(T_2) \cdot \sum \cdot \cdots \\
&= p(\delta_{ij}) \cdot \overline{\alpha_{ij}}(W_d)
\end{aligned}
$$

ここで, $\overline{\alpha_{ij}}(W_d)$ は, 変数の積$\overline{\alpha_{ij}}$における $X_1, X_2, \cdots$ にそれぞれ $[W_d]_1, [W_d]_2 \cdots$ を代入した値である.

よって,

$$
\begin{aligned}
& [W_{d+1}]_i \\
& \quad = P_r(\mathcal{T}_{d+1}(X_i)) = \sum_{j=1}^{m_i} P_r(\mathcal{T}_{d+1}(X_i \to \alpha_{ij})) \\
& \quad = \sum_{j=1}^{m_i} p(\delta_{ij}) \cdot \overline{\alpha_{ij}}(W_d) = [F(W_d)]_i
\end{aligned}
$$

となり, i が成立する.

X_i からの構文木で深さ 0 の木は存在しないので, $W_0 = 0$. このことと i を用いて, d に関する数学的帰納法により, ii が証明される. □

【 定理 1 】 次の i, ii が成立する.

i. $W = \lim_{d\to\infty} F^d(0)$ に対し, 次の式が成立する.

$$0 \leq W = F(W).$$

また, $0 \leq A = F(A)$ ならば, $W \leq A$ である. (以後, W を特性方程式 (9) の **正値最小解**と呼ぶ.)

ii. 正値最小解 W の 第 i 成分 $[W]_i$ は, X_i からのすべての構文木の生起確率の総和に等しい.

証明. 補題 4.1 の iii より, $W = \lim_{d\to\infty} F^d(0)$ が存在する. また, $[F(\boldsymbol{X})]_i$ が 連続関数であることより,

$$
\begin{aligned}
F(W) &= F(\lim_{d\to\infty} F^d(0)) = \lim_{d\to\infty} F(F^d(0)) \\
&= \lim_{d\to\infty} F^{d+1}(0) = W.
\end{aligned}
$$

このことと, 補題 4.1 の ii より, $\leq$ に関して 0 以上の特性方程式の解の中で, W は $\leq$ に関して最小な解 (正値最小解) となり, i が成立する.

$\mathcal{T}_1(X_i), \mathcal{T}_2(X_i), \cdots$ は集合の包含関係に関して単調に増大すること, $\lim_{d\to\infty} \mathcal{T}_d(X_i)$ は X_i からの構文木の全体集合であること, および i から, ii が成立する. □

例 1 の場合, 特性方程式 $\boldsymbol{X} = F(\boldsymbol{X})$ の解は 2 つ存在し,

$$\boldsymbol{X} = \begin{bmatrix} -2 \\ -2 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix}.$$

よって, 正値最小解は$\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ だから, G は健全である.

例 2. $G = (\Sigma, \{X_1\}, P, X_1, p)$,

$P = \{X_1 \underset{\frac{2}{3}}{\longrightarrow} X_1X_1, X_1 \underset{\frac{1}{3}}{\longrightarrow} a\}$.

特性方程式 $X_1 = \frac{2}{3}X_1^2 + \frac{1}{3}$ の解は, $\frac{1}{2}, 1$. したがって, 正値最小解は$\frac{1}{2}$. このことは, すべての構文木の生起確率の総和が $\frac{1}{2}$ であることを意味しており, G は不健全である. この例における p を少し変えて, $P = \{X_1 \underset{\frac{1}{2}}{\longrightarrow} X_1X_1, X_1 \underset{\frac{1}{2}}{\longrightarrow} a\}$ とすると, 正値最小解は 1 となり, G は健全になる. □

定理 1 は Booth & Thompson の条件と比較して, 次の点で優れている.

i 例 1 に示されるように, Booth & Thompson の条件は文法が健全であるための十分条件にすぎない. 一方, 定理 1, i は, 特性方程式の正値最小解が **1** であることが, 文法が健全であるための必要十分条件であることを示している.

ii Booth & Thompson は, すべての構文木の生起確率の総和 ($= [W]_1$) を計算する方法を与えていないが, 定理 1 はこれを与えている. 特に, $F(X)$ の次数や変数の個数が多くて特性方程式の解を求めることが難しい場合でも, ii を用いて, 十分な精度で Wを計算することができる.

5 書換え規則の適用頻度のモーメント

【定義 6】 木 T において，書換え規則 $\delta \in P$ が適用された回数を $c(\delta,T)$ と記し，n 次元ベクトル $\overline{C}(\delta),\overline{C^2}(\delta)$，を次のように定義する．

$$[\overline{C}(\delta)]_i \stackrel{\mathrm{def}}{=} \sum_{T\in\mathcal{T}(X_i)} c(\delta,T)P_r(T),$$

$$[\overline{C^2}(\delta)]_i \stackrel{\mathrm{def}}{=} \sum_{T\in\mathcal{T}(X_i)} c^2(\delta,T)P_r(T).$$

□

$[\overline{C}(\delta)]_i,[\overline{C^2}(\delta)]_i$ は，それぞれ，X_i からの構文木において $\delta(\in P)$ が適用される回数の一次モーメント（平均値），二次モーメントである．

木 $T\in\mathcal{T}(\delta_{ij})$ において，root X_i の子 $X_{k(\alpha_{ij})}$ からの部分木を $T_k \in \mathcal{T}(X_{k(\alpha_{ij})})$ で表すと，

$$c(\delta,T) = \begin{cases} c(\delta,T_1)+c(\delta,T_2)+\cdots & (\delta\neq\delta_{ij}\text{のとき}) \\ 1+c(\delta,T_1)+c(\delta,T_2)+\cdots & (\delta=\delta_{ij}\text{のとき}) \end{cases},$$

$$P_r(T) = p(\delta_{ij})\Pi_k P_r(T_k).$$

よって，$\delta\neq\delta_{ij}$ のとき

$$\sum_{T\in\mathcal{T}(X_i\to\alpha_{ij})} c(\delta,T)P_r(T)$$

$$= p(\delta_{ij})\sum_k \frac{\overline{\alpha_{ij}}(W)}{[W]_{k(\alpha_{ij})}}[\overline{C}(\delta)]_{k(\alpha_{ij})}$$

$$= p(\delta_{ij})\sum_{l=1}^{n}\frac{\partial\overline{\alpha_{ij}}}{\partial X_l}(W)[\overline{C}(\delta)]_l$$

$$= p(\delta_{ij})\left[\frac{\partial\overline{\alpha_{ij}}}{\partial X}(W)\circ\overline{C}(\delta)\right]. \quad \cdots ①$$

ここで，$k(\alpha_{ij})$ は α_{ij} における左から k 番目の非終端記号の index，$\frac{\partial\overline{\alpha_{ij}}}{\partial X_l}(W)$ は，$\overline{\alpha_{ij}}$ の一次偏微分 $\frac{\partial\overline{\alpha_{ij}}}{\partial X_l}$ において $X_1,\cdots,X_n$ に $[W]_1,\cdots,[W]_n$ を代入した値であり，$\frac{\partial\overline{\alpha_{ij}}}{\partial X}(W)$ は $\frac{\partial\overline{\alpha_{ij}}}{\partial X_l}(W)$ を第 l 成分とする n 次元ベクトル，$\circ$ はベクトルの内積演算である．

また，$\delta=\delta_{ij}$ の場合も同様にして，②を得る．

$$\sum_{T\in\mathcal{T}(X_i\to\alpha_{ij})} c(\delta,T)P_r(T) =$$

$$p(\delta_{ij})\cdot\left[\frac{\partial\overline{\alpha_{ij}}}{\partial X}(W)\circ\overline{C}(\delta)+\overline{\alpha_{ij}}(W)\right]\cdots ②$$

①②より，③が成立する．

$$[\overline{C}(\delta_{ij})]_l =$$

$$\begin{cases} \frac{\partial f_l}{\partial X}(W)\circ\overline{C}(\delta_{ij}) & (l\neq i\text{のとき}) \\ \frac{\partial f_l}{\partial X}(W)\circ\overline{C}(\delta_{ij})+p(\delta_{ij})\cdot\overline{\alpha_{ij}}(W) & (l=i\text{のとき}) \end{cases}$$

$$\cdots ③$$

【定義 7】 $\frac{\partial f_l}{\partial X_i}(W)$ を第 l 成分とする n 次元ベクトルを $\frac{\partial F}{\partial X_i}(W)$ と記し，$\frac{\partial F}{\partial X_i}(W)$ を第 i 列とするような $n\times n$ 行列を J と記す．

$$J \stackrel{\mathrm{def}}{=} \left(\frac{\partial F}{\partial X_1}(W),\cdots,\frac{\partial F}{\partial X_n}(W)\right)$$

また，第 i 成分が 1 で他の成分が 0 のベクトルを e_i と記す． □

③より，次の定理を得る．

【定理 2】

i $\overline{C}(\delta_{ij}) = J\overline{C}(\delta_{ij})+p(\delta_{ij})\overline{\alpha_{ij}}(W)e_i$

ii $J^* \stackrel{\mathrm{def}}{=} J^0+J^1+J^2+\cdots$ が存在するならば，

$$\overline{C}(\delta_{ij}) = p(\delta_{ij})\overline{\alpha_{ij}}(W)\cdot J^*e_i$$

証明． i は ③より自明である．また，i の式の両辺に，左から J^* を乗ずることにより，ii を得る． □

木 $T\in\mathcal{T}(X_i\to\alpha_{ij})$ において，root X_i の子 $X_{k(\alpha_{ij})}$ からの部分木を $T_k\in\mathcal{T}(X_{k(\alpha_{ij})})$ で表すと，

$\delta\neq\delta_{ij}$ ならば，

$$c^2(\delta,T) = \left(\sum_k c(\delta,T_k)\right)^2$$

$$= \sum_k c^2(\delta,T_k)+\sum_k\sum_{k'\neq k}c(\delta,T_k)c(\delta,T_{k'})$$

$\delta=\delta_{ij}$ ならば，

$$c^2(\delta,T) = \left(1+\sum_k c(\delta,T_k)\right)^2$$

$$= \sum_k c^2(\delta,T_k)+\sum_k\sum_{k'\neq k}c(\delta,T_k)c(\delta,T_{k'})$$

$$+2\sum_k c(\delta,T_k)+1.$$

また，$P_r(T) = p(\delta_{ij})\cdot\Pi P_r(T_k)$.

よって，$\delta \neq \delta_{ij}$ のとき，

$$\sum_{T\in T(X_i\to\alpha_{ij})} c^2(\delta,T)P_r(T)$$

$$= p(\delta_{ij})\sum_{k} \frac{\overline{\alpha_{ij}}(W)}{[W]_{k(\alpha_{ij})}}[\overline{C^2}(\delta)]_{k(\alpha_{ij})}$$

$$+p(\delta_{ij})\sum_{k}\sum_{k'\neq k} \frac{\overline{\alpha_{ij}}(W)}{[W]_{k(\alpha_{ij})}[W]_{k'(\alpha_{ij})}}[\overline{C}(\delta)]_{k(\alpha_{ij})}[\overline{C}(\delta)]_{k'(\alpha_{ij})}$$

$$= p(\delta_{ij})\sum_{l=1}^{n} \frac{\partial\overline{\alpha_{ij}}}{\partial X_l}(W)[\overline{C^2}(\delta)]_l$$

$$+p(\delta_{ij})\sum_{l=1}^{n}\sum_{m=1}^{n} \frac{\partial^2\overline{\alpha_{ij}}}{\partial X_l\partial X_m}(W)[\overline{C}(\delta)]_l[\overline{C}(\delta)]_m \quad \cdots ④$$

$\delta = \delta_{ij}$ のとき，

$$\sum_{T\in T(X_i\to\alpha_{ij})} c^2(\delta,T)P_r(T)$$

$$= p(\delta_{ij})\sum_{l=1}^{n} \frac{\partial\overline{\alpha_{ij}}}{\partial X_l}(W)[\overline{C^2}(\delta)]_l$$

$$+p(\delta_{ij})\sum_{l=1}^{n}\sum_{m=1}^{n} \frac{\partial^2\overline{\alpha_{ij}}}{\partial X_l\partial X_m}(W)[\overline{C}(\delta)]_l[\overline{C}(\delta)]_m$$

$$+2p(\delta_{ij})\sum_{l=1}^{n} \frac{\partial\overline{\alpha_{ij}}}{\partial X_l}(W)[\overline{C}(\delta)]_l + p(\delta_{ij})\overline{\alpha_{ij}}(W) \quad \cdots ⑤$$

【定義 8】 $\frac{\partial f_i}{\partial X_l \partial X_m}(W)$ を (l,m) 要素とする $n\times n$ 行列を H_iと記す. □

④⑤より，⑥を得る.

$$[\overline{C^2}(\delta_{ij})]_l =$$

$$\begin{cases} \frac{\partial f_l}{\partial X}(W)\circ\overline{C^2}(\delta_{ij}) + H_l\overline{C}(\delta_{ij})\circ\overline{C}(\delta_{ij}) & \cdots (l\neq i) \\ \frac{\partial f_l}{\partial X}(W)\circ\overline{C^2}(\delta_{ij}) + H_l\overline{C}(\delta_{ij})\circ\overline{C}(\delta_{ij}) & \\ \quad +2p(\delta_{ij})\frac{\partial\overline{\alpha_{ij}}}{\partial X}(W)\circ\overline{C}(\delta) + p(\delta_{ij})\overline{\alpha_{ij}}(W) & \cdots (l=i) \end{cases} \quad \cdots ⑥$$

【定理 3】

i $\quad \overline{C^2}(\delta_{ij}) = J\overline{C^2}(\delta_{ij}) + e(\delta_{ij})$

$$\text{ただし，} e(\delta_{ij}) = \begin{bmatrix} H_1\overline{C}(\delta_{ij})\circ\overline{C}(\delta_{ij}) \\ \vdots \\ H_n\overline{C}(\delta_{ij})\circ\overline{C}(\delta_{ij}) \end{bmatrix}$$

$$+p(\delta_{ij})\left(2\frac{\partial\overline{\alpha_{ij}}}{\partial X}(W)\circ\overline{C}(\delta) + \overline{\alpha_{ij}}(W)\right)e_i.$$

ii J^* が存在するならば，$\quad \overline{C^2}(\delta_{ij}) = J^* e(\delta_{ij})$

□

【定理 4】

i 行列 J^*の各要素が有限な値，すなわち，J^* が存在することと，任意の$\delta \in P$に対し，δが構文木において適用される回数の平均値と分散が共に有限であることは同値である.

ii J^* が存在するとき，

$$[\overline{C}(\delta_{ij})]_1 = p(\delta_{ij})[J^*]_{1i}\overline{\alpha_{ij}}(W)$$

証明　構文木における $c(\delta,T)$ の期待値は $[\overline{C}(\delta)]_1$であり，分散は $[\overline{C}(\delta)]_1^2 - [\overline{C^2}(\delta)]_1$ である．したがって，J^* が存在すれば，任意の$\delta \in P$に対し，$c(\delta,T)$ の平均値と分散は存在する.

いま，$J^* = J^0 + J^1 + J^2 + \cdots$ の (i,j) 要素 $[J^*]_{ij}$が∞ であるとする．定理 2,i の式において，左辺の$\overline{C}(\delta_{ij})$ に $\overline{C}(\delta_{ij}) = J\overline{C}(\delta_{ij}) + p(\delta_{ij})\overline{\alpha_{ij}}(W)e_i$を k回代入することにより，

$$\overline{C}(\delta_{jl}) = J^{k+1}\overline{C}(\delta_{jl}) + p(\delta_{jl})\overline{\alpha_{jl}}(W)(J^0 + J^1 + \cdots + J^k)e_i$$

$$\therefore \quad [\overline{C}(\delta_{jl})]_i \geq (\delta_{jl})\overline{\alpha_{jl}}(W)[J^0 + J^1 + \cdots + J^k]_{ij}$$

$[J^0+J^1+\cdots+J^k]_{ij}$は，$k$ の増加にともなって$\infty(=[J^*]_{ij})$になるから，$[\overline{C}(\delta_{jl})]_i = \infty$ となる．また，定理 3,i の式に同様の手法を用いて $[\overline{C^2}(\delta_{jl})]_i = \infty$ が導かれる．本論文で扱う文法 G は well formed (定義 1 参照) であるから，$[\overline{C}(\delta_{jl})]_i = \infty$ ならば，$[\overline{C}(\delta_{jl})]_1$ も ∞ になることが示されるが，紙面の都合上証明は割愛する．よって，i が成立する.

ii が成立することは，定理 2, ii より自明である. □

6 一致推定

G が健全 ($W = \mathbf{1}$，したがって$\overline{\alpha_{jl}}(W) = 1$) で，かつ J^* が存在する場合には，定理 5,ii より，任意の $\delta_{ij} \in P$ に対して，X_iからの構文木 Tについての $c(\delta_{ij},T)$ の平均値は

$$[\overline{C}(\delta_{ij})]_1 = p(\delta_{ij})[J^*]_{1i} \quad \cdots ①$$

であり，かつ $c(\delta_{ij},T)$ の分散は次式のようになり，有限である.

$$\sum_{T\in T(X_1)} (c(\delta_{ij},T) - [\overline{C}(\delta_{ij})]_1)^2 P(T)$$

$= [\overline{C^2}(\delta_{ij})]_1 - [\overline{C}(\delta_{ij})]_1^2 \quad \cdots ②$

①から分かるように，$[J^*]_{1i}$ は jに無関係だから，X_1 からの構文木における $\delta_{i1}, \delta_{i2} \cdots, \delta_{im_i}$の平均適用回数の比は $p(\delta_{i1}), p(\delta_{i2}), \cdots, p(\delta_{im_i})$ になる．したがって，N 個の構文木の標本列 $\mathcal{F}_N = (T_1, T_2, \cdots, T_N)$ に対し，次の定義のように $p(\delta_{ij})$ の推定式 $\widehat{p}(\delta_{ij}\mathcal{F}_N)$ を構成すれば良い．

【定義 9】 推定式

$$\widehat{p}(\delta_{ij}, \mathcal{F}_N) \stackrel{\text{def}}{=} \frac{\sum_{k=1}^{N} c(\delta_{ij}, \mathcal{F}_N(k))}{\sum_{l=1}^{m_i}\sum_{k=1}^{N} c(\delta_{il}, \mathcal{F}_N(k))}$$

ただし，$\mathcal{F}_N(k)$ は標本列 $\mathcal{F}_N$における k番目の構文木である． □

【定義 10】 標本列 $\mathcal{F}_N$の実数値関数 $f(\mathcal{F}_N)$ が実数 a に<u>確率収束</u>するとは，次のことである．すなわち，任意の $\varepsilon > 0$ に対して N_0が存在し，$N > N_0$ に対して，

$$P_r\{|f(\mathcal{F}_N) - a| < \varepsilon\} > 1 - \varepsilon.$$

□

【定理 5】 一致性

G が健全で，かつ J^*が存在する場合には，定義9の推定式 $\widehat{p}(\delta_{ij}, \mathcal{F}_N)$ は一致推定である．

証明．

$$A(i,j,\mathcal{F}_N) = \sum_{k=1}^{N} c(\delta_{ij}, \mathcal{F}_N(k))/N, \quad \overline{A}(i,j) = [\overline{C}(\delta_{ij})]_1,$$

$$B(i,\mathcal{F}_N) = \sum_{l=1}^{m_i}\sum_{k=1}^{N} c(\delta_{il}, \mathcal{F}_N(k))/N,$$

$$\overline{B}(i) = \sum_{l=1}^{m_i} [\overline{C}(\delta_{il})]_1$$

と置く．

$c(\delta_{il}, T)$ の平均値と分散が存在するから，統計学における大数の法則より，任意の i,jに対して $A(i,j,\mathcal{F}_N)$ は $A(i,j)$ に確率収束する．また，確率収束する2つの関数の和は，また確率収束するので，$B(i,\mathcal{F}_N) = \sum_{l=1}^{m_i} A(i,,\mathcal{F}_N)$ は$\overline{B}(i)$ に確率収束する．よって

$\forall\varepsilon(>0)\exists N_0 \ \forall N(>N_0)$

$$P_r\{|A(i,j,\mathcal{F}_N) - \overline{A}(i,j)| < \varepsilon\} > 1 - \varepsilon,$$

$$P_r\{|B(i,\mathcal{F}_N) - \overline{B}(i)| < \varepsilon\} > 1 - \varepsilon.$$

$|A(i,j,\mathcal{F}_N) - \overline{A}(i,j)| < \varepsilon$, かつ $|B(i,\mathcal{F}_N) - \overline{B}(i)| < \varepsilon$ である $\mathcal{F}_N$に対して，

$$\frac{\overline{A}(i,j) - \varepsilon}{\overline{B}(i) + \varepsilon} < \frac{A(i,j,\mathcal{F}_N)}{B(i,\mathcal{F}_N)} < \frac{\overline{A}(i,j) + \varepsilon}{\overline{B}(i) - \varepsilon}$$

だから，

$$\left|\frac{A(i,j,\mathcal{F}_N)}{B(i,\mathcal{F}_N)} - \frac{\overline{A}(i,j)}{\overline{B}(i)}\right| < \frac{\varepsilon(\overline{A}(i,j) + \overline{B}(i))}{\overline{B}(i)(\overline{B}(i) - \varepsilon)}.$$

よって，$\varepsilon' = max\left(2\varepsilon, \frac{\varepsilon(\overline{A}(i,j)+\overline{B}(i))}{\overline{B}(i)(\overline{B}(i)-\varepsilon)}\right)$ と置くと，

$$P_r\left\{\left|\frac{A(i,j,\mathcal{F}_N)}{B(i,\mathcal{F}_N)} - \frac{\overline{A}(i,j)}{\overline{B}(i)}\right| < \varepsilon'\right\} > 1 - \varepsilon'.$$

となり，

$\widehat{p}(\delta_{ij}, \mathcal{F}_N) = \frac{A(i,j,\mathcal{F}_N)}{B(i,\mathcal{F}_N)}$ は

$$\frac{\overline{A}(i,j)}{\overline{B}(i)} = \frac{p(\delta_{ij})}{\sum_{l=1}^{m_i} p(\delta_{il})} = p(\delta_{ij})$$

に確率収束する． □

7 おわりに

確率文脈自由文法のパラメタ推定法を議論した．文法が健全でない場合には，精度の高いパラメタ推定が難しいことが，定理 4,ii の式よりうかがわれる．文法が健全で J^*が存在する場合に，一致推定式を与えた．文法が健全であって，かつ J^*が存在しない場合の一致推定についても見通しを得ており，近く発表する．

参考文献

[1] T.L.Booth, R.A.Thompson : *Applying Probability Measures to Abstract Languages*, IEEE Trans. Comput. C-22, pp.442-450 (1973)

[2] K.Lari, S.J.Young : *The Estimation of Stochastic Context-Free Grammars Using the Inside-Outside Algorithm*, Computer Speech and Language, Vol.4, pp.35-36 (1990)

[3] 日高　達：確率文法，情報処理学会誌，36 巻第 2 号，pp.169-176(1995)