数理解析研究所講究録
第 1849 巻 2013 年 8-11

森および連結全域部分グラフの乱択近似数え上げ
Randomized Approximate Counting of Forests and Connected Spanning Subgraphs

三原 勇治 Yuji Mihara　山内 由紀子 Yukiko Yamauchi　来嶋 秀治 Shuji Kijima　山下 雅史 Masafumi Yamashita

九州大学
Kyushu University

1 はじめに

組合せ構造の近似数え上げと（一様）ランダム生成の間には，自己帰着性（self-reducibility）を通じて密接に関係することが知られている．基本的な組み合わせ構造の一つである全域木のランダム生成には [2] を初め，様々な多項式時間アルゴリズムが知られている．一方，全域木の一般化にあたる森や連結全域部分グラフの数え上げ問題は#P 完全であることが知られ，多項式時間近似アルゴリズムの存在性は重要な未解決問題である．

マルコフ連鎖モンテカルロ（MCMC）法は，森および連結全域部分グラフの多項式時間一様ランダム生成法として有望である．MCMC の計算効率性について，設計したマルコフ連鎖が定常分布に十分収束するまでに必要な遷移回数（mixing time:混交時間）が解析の焦点となる．Fehrenbach と Rüschendorf [2] は全域木上のマルコフ連鎖に対し，多品種流を用いた解析により，混交時間が頂点数の多項式で抑えられることを示した．Fehrenbach と Rüschendorf の解析では，Cordovil と Moreira [1] が示した全域木に対する bases-cobases グラフの連結性が重要な役割を果たしている．本研究では，森および連結全域部分グラフに対して Cordovil と Moreira [1] の結果を拡張する．この結果を用いて，辺数が全域木よりも 1 つ少ない森上のマルコフ連鎖に対し，[2] と同様の解析を行い，混交時間が頂点数の多項式で抑えられることを示す．

2 準備

2.1 マルコフ連鎖の定常分布

有限の状態空間 Ω と遷移確率行列 P によって定まる，マルコフ連鎖を考える．マルコフ連鎖の定常分布とは，$\pi P = \pi$ を満たすような Ω 上の確率分布 π である．マルコフ連鎖が以下の二つの性質を満たすとき，エルゴード的であると呼び，唯一の定常分布に収束することが知られている．

1. 既約，すなわち任意の状態 $x, y \in \Omega$ について $P^t(x,y) > 0$ となる t が存在する．
2. 非周期，すなわち任意の状態 $x \in \Omega$ について $\gcd\{t \mid P^t(x,x) > 0\} = 1$ である．ここで gcd は最大公約数のことである．

マルコフ連鎖モンテカルロ (MCMC) 法は，サンプリングにマルコフ連鎖を用いる手法である．一般的なサンプリング方法は，まずサンプリングを行いたい目標分布を定常分布として持ち，かつそれに収束するようなマルコフ連鎖を設計する．次に，ある初期状態から遷移を開始し，十分遷移させた後の状態を出力することでサンプルを得る．

2.2 マルコフ連鎖の混交時間

混交時間 (mixing time) とは，マルコフ連鎖が定常分布に収束するまでに必要な遷移回数である．ま

ず，定常分布との距離の指標として時刻 t における全変動距離 (total variation distance) を

$$d_{TV}(P^t,\pi):=\max_{x\in\Omega}\frac{1}{2}\sum_{y\in\Omega}|P^t(x,y)-\pi(y)|$$

と定義する．この全変動距離を用いて，任意の $\epsilon>0$ に対して混交時間 $\tau(\epsilon)$ は以下のように定義される．

$$\tau(\epsilon):=\min\{t\,|\,\forall t'\geq t,\ d_{TV}(P^{t'},\pi)\leq\epsilon\}$$

2.3 Bases-cobases グラフ

E を有限集合とし，$\mathcal{F}\subseteq 2^E$ が以下の 3 つの性質を持つとき，E と $\mathcal{F}$ の組 $(E,\mathcal{F})$ を**マトロイド**と呼ぶ．

1. $\emptyset\in\mathcal{F}$
2. $F_1\subseteq F_2\in\mathcal{F}\Rightarrow F_1\in\mathcal{F}$
3. F_1，$F_2\in\mathcal{F}$，$|F_1|<|F_2|\Rightarrow\exists e\in F_2\setminus F_1:F_1\cup\{e\}\in\mathcal{F}$

マトロイド $M=(E,\mathcal{F})$ に対し，E を台集合，$\mathcal{F}$ を独立集合族，各 $F\in\mathcal{F}$ を独立集合と呼ぶ．また，極大な独立集合を基と呼ぶ．

マトロイド M に対し，その台集合 E がある 2 つの排反な基の和集合となっているとき，M を**ブロックマトロイド**と呼ぶ．ブロックマトロイド M に対し，M の基 B で，その補基 $E\setminus B$ も M の基であるようなものを頂点とし，要素がちょうど一つ異なる基の組 (B,B') を辺とするグラフを M の **bases-cobases グラフ**と呼ぶ．

2.4 多品種流による混交時間の解析

Ω 上のマルコフ連鎖 M に対し，$\mathcal{P}_{xy}$ を M の状態遷移図における状態 x から状態 y への全ての単純有向道の集合とする．$x,y\in\Omega$ に対し，$f_{xy}:\mathcal{P}_{xy}\to\Re_0^+$ を $\sum_{p\in\mathcal{P}_{xy}}f_{xy}(p)=1$ を満たす関数とした時，$F=\{f_{xy};x,y\in\Omega\}$ を**多品種流**と呼ぶ．F の**混雑度** $\rho(F)$ を

$$\rho(F)=\max_{P(v,w)>0}\frac{1}{\pi(v)P(v,w)}\sum_{p_{xy}\ni(v,w)}\pi(x)\pi(y)f_{xy}(p_{xy})|p_{xy}|$$

と定義する．但し π は，マルコフ連鎖 M の定常分布とする．この混雑度を用いて，混交時間に対して次のような上界が得られる．

定理 1 [3]
Ω 上のマルコフ連鎖が

(1) エルゴード的で，π を定常分布に持つ

(2) $\forall x,y\in\Omega,\ \pi(x)P(x,y)=\pi(y)P(y,x)$

(3) $\forall x\in\Omega,\ P(x,x)\geq\frac{1}{2}$

を満たす時，その混交時間 $\tau(\epsilon)$ と任意の多品種流 F に対し，

$$\tau(\epsilon)\leq\rho(F)(\log\hat{\pi}^{-1}+\log\epsilon^{-1})$$

が成立する．但し，$\hat{\pi}=\min_{x\in\Omega}\pi(x)$ である．

3 森/連結全域部分グラフに対する bases-cobases グラフの連結性

Cordovil と Moreira [1] の拡張を与える．グラフ $G=(V,E)$ に対して，E がサイズ r の排反な二つの森の和である時，G 中のサイズ r の森全体を基とするブロックマトロイドが存在する．

定理 2 任意の $r\in\mathbb{N}$ に対し，M をサイズ r の森を基とするブロックマトロイドとする．M の bases-cobases グラフにおいて，相補的な基に対応する任意の頂点対は長さ r の道で結ばれている．

同様に，E があるサイズ r の排反な二つの連結全域部分グラフの和である時，G 中のサイズ r の連結全域部分グラフ全体を基とするブロックマトロイドが存在する．

定理 3 任意の $r\in\mathbb{N}$ に対し，M をサイズ r の連結全域部分グラフを基とするブロックマトロイドとする．M の bases-cobases グラフにおいて，相補的な基に対応する任意の頂点対は長さ r の道で結ばれている．

4 辺数$n-2$の森上のマルコフ連鎖に対する混交時間の解析

無向グラフ$G=(V,E)$に対し，$|V|=n$，$|E|=m$とする．$\Omega=\{F\subseteq E\mid F$は森,$|F|=n-2\}$を状態空間とするマルコフ連鎖$M$は，現状態を$X\in\Omega$とした時，以下のように次状態を決定する．

1. $e\in X$と$f\in E$をそれぞれ独立かつ一様ランダムに選ぶ．
2. $Y=(X\setminus\{e\})\cup\{f\}$とする．
3. (a) $Y\in\Omega$の場合，次状態を確率$\frac{1}{2}$でY，確率$\frac{1}{2}$でXとする．

   (b) $Y\notin\Omega$の場合，次状態をXとする．

定理 4 マルコフ連鎖Mは一様分布を唯一の定常分布にもち，かつそれに収束する．

また混交時間に関して以下の定理を得る．

定理 5 任意の$\epsilon\in(0,1)$に対し，グラフ$G=(V,E)$に対するマルコフ連鎖Mの混交時間$\tau(\epsilon)$に対して

$$\tau(\epsilon)\leq 2n^4m(n\ln m+\ln\epsilon^{-1})$$

が成立する．ただし，$n:=|V|$，$m:=|E|$である．

マルコフ連鎖Mの混交時間を解析するために，多品種流を構成する．すなわち，任意の$X,Y\in\Omega$に対して，関数$f_{XY}:\mathcal{P}_{XY}\to\Re_0^+$を定義する．ここで，$X$と$Y$を重ね合わせ，共通の辺を縮約してできるグラフ$H=(V,X\cup Y)/(X\cap Y)$を考える．$H$上では$X\setminus Y$と$Y\setminus X$は排反な森であり，$|X\setminus Y|=|Y\setminus X|$である．したがって，$\ell:=|X\setminus Y|=|Y\setminus X|$とすると，$H$中のサイズ$\ell$の森を基とするブロックマトロイドが存在する．$X$から$Y$への道の内，このブロックマトロイドのbases-cobasesグラフ上を移動するような道pに対してのみ$f_{XY}(p)>0$とする．すなわち，$p=(B_i)_{0\leq i\leq\ell}$とすると，任意の$i$に対して，$H$上での$B_i$の補グラフも森である道に対して$f_{XY}(p)>0$とする．

多品種流の構成法について以下述べる．$X\cup A$と$Y\cup A$が共に森であるような任意の$A\subset E\setminus(X\cup Y)$に対し，$\sum_{p\in\mathcal{P}_{XY}}f_{XY}^A(p)=1$を満たす関数$f_{XY}^A:\mathcal{P}_{XY}\to\Re_0^+$を定義し，$f_{XY}(p)=f_{XY}^{\emptyset}(p)$とする．また$H_A=(V,X\cup Y)/((X\cap Y)\cup A)$とする．$H_A$は$H$に対して$A$の辺を縮約してできるグラフであり，$H_A$でも$X\setminus Y$と$Y\setminus X$は排反な森である．実際に$X\cup A$と$Y\cup A$が共に森であるような任意の$A\subset E\setminus(X\cup Y)$に対して$f_{XY}^A$を定義する際は，まず$H_A$の各連結成分において，$X\setminus Y$と$Y\setminus X$の辺が排反な二つの全域木を構成する$A$に対して$f_{XY}^A$を定義する．

ℓに関して再帰的にf_{XY}^Aを定義する．$\ell=1$の場合，$P(X,Y)>0$である．この場合，XからYへの遷移だけからなる道$p=(B_i)_{0\leq i\leq 1}$(すなわち，$B_0=X$，$B_1=Y$)に対して$f_{XY}^A(p)=1$とし，それ以外の道$p'\in\mathcal{P}_{XY}\setminus p$に対しては$f_{XY}^A(p')=0$とする．

$\frac{1}{2}|X\oplus Y|<\ell$の場合，$X\cup A$と$Y\cup A$が共に森であるような任意の$A\subset E\setminus(X\cup Y)$に対し，既に$f_{XY}^A$は定義されていると仮定する．$\frac{1}{2}|X\oplus Y|=\ell$の場合を考える．$d_{\min}$を$H_A$上の孤立点以外の頂点の最小次数とし，$D$を次数が$d_{\min}$である頂点の集合とする．各$v\in D$に対し，以下の(1)，(2)を満たすような$X',Y'\in\Omega$と$A'\subset E$の組$(X',Y',A')$の集合$S_v$を選ぶ．

(1) $\frac{1}{2}|X'\oplus Y'|=\ell-1$

(2) $X'\cup(A\cup A'),Y'\cup(A\cup A')$は共に森

再帰の仮定より，任意の$(X',Y',A')\in S_v$に対し，$f_{X'Y'}^{A\cup A'}$は既に定義されている．さらに，$f_{X'Y'}^{A\cup A'}(p')>0$である各道$p'\in\mathcal{P}_{X'Y'}$に対し，それを拡張して$p\in\mathcal{P}_{XY}$を作り，

$$f_{XY}^A(p):=\frac{1}{|D|}\frac{1}{|S_v|}f_{X'Y'}^{A\cup A'}(p')$$

とする．

次に，$v\in D$に対して，$X',Y'\in\Omega$と$A'\subset E$の組の集合S_vの選び方と，$\mathcal{P}_{X'Y'}$の道から$\mathcal{P}_{XY}$の道の作り方を示す．vに接続している辺の集合をC_vとする．XとYは辺数$n-2$の森であるので，握手補題より$d_{\min}<4$である．$d_{\min}=1,2,3$のそれぞれの場合を考える．

<u>Case 1.</u> $d_{\min}=1$の場合，$C_v=\{a\}$とし，一般性を失うことなく$a\in X$とする．このとき，$C=\{d\in Y\mid X\cup\{d\}$は閉路をもたない(全域木)$\}$とする．各

$d \in C$ に対し，$X_v^d = (X \setminus \{a\}) \cup \{d\}$，$Y_v^d = Y$ とする．$\frac{1}{2}|X_v^d \oplus Y_v^d| = \ell - 1$ より，X_v^d，Y_v^d の組は再帰の仮定を満たす．さらに，$X_v^d \cup \{a\}$ と $Y_v^d \cup \{a\}$ は共に森である．この場合，$S_v = \{(X_v^d, Y_v^d, \{a\}) | d \in C\}$ とする．道 $p' = (B'_i)_{0 \leq i < \ell} \in \mathcal{P}_{X_v^d Y_v^d}$ に対して，最初に X から X_v^d への遷移を加えることで道 $p = (B_i)_{0 \leq i \leq \ell} \in \mathcal{P}_{XY}$ にできる．
Case 2. $d_{\min} = 2$ の場合，$C_v = \{a, b\}$ とする．まず $C_v \subseteq X$ または $C_v \subseteq Y$ ならば，一般性を失うことなく $C_v \subseteq X$ とする．このとき，$C = \{d \in Y \mid X \cup \{d\}$ は閉路をもたない (全域木)$\}$ とする．各 $e \in C_v, d \in C$ に対し，$X_v^{ed} = (X \setminus \{e\}) \cup \{d\}$，$Y_v^{ed} = Y$ とする．$\frac{1}{2}|X_v^{ed} \oplus Y_v^{ed}| = \ell - 1$ より，X_v^{ed}，Y_v^{ed} の組は再帰の仮定を満たす．さらに，$X_v^{ed} \cup \{e\}$ と $Y_v^{ed} \cup \{e\}$ は共に森である．この場合，$S_v = \{(X_v^{ed}, Y_v^{ed}, \{e\}) \mid e \in C_v,\ d \in C\}$ とする．道 $p' = (B'_i)_{0 \leq i < \ell} \in \mathcal{P}_{X_v^{ed} Y_v^{ed}}$ に対して，最初に X から X_v^{ed} への遷移を加えることで道 $p = (B_i)_{0 \leq i \leq \ell} \in \mathcal{P}_{XY}$ にできる．$C_v \subseteq X$ でも $C_v \subseteq Y$ でもない場合，一般性を失うことなく $a \in X$，$b \in Y$ とする．$X_v = (X \setminus \{a\}) \cup \{b\}$，$Y_v = Y$ とすることで，X_v，Y_v の組は再帰の仮定を満たす．この場合，$S_v = \{(X_v, Y_v, \emptyset)\}$ とする．道 $p' = (B'_i)_{0 \leq i < \ell} \in \mathcal{P}_{X_v Y_v}$ に対して，最初に X から X_v への遷移を加えることで道 $p = (B_i)_{0 \leq i \leq \ell} \in \mathcal{P}_{XY}$ にできる．
Case 3. $d_{\min} = 3$ の場合，$C_v = \{a, b, c\}$ とする．$C_v \subseteq X$ または $C_v \subseteq Y$ ならば，$d_{\min} = 2$ の場合と同様である．そうではない場合，一般性を失うことなく $a, b \in X$，$c \in Y$ とする．各 $e \in \{a, b\}$ に対し，$X_v^e = X$，$Y_v^e = (Y \setminus \{c\}) \cup \{e\}$ とする．Y_v^e が森ならば，X_v^e，Y_v^e の組は再帰の仮定を満たす．この場合，$S_v = \{(X_v^e, Y_v^e, \emptyset) \mid e \in \{a, b\}, Y_v^e$ は森$\}$ とする．以降は一般性を失うことなく $e = a$ とする．道 $p' = (B'_i)_{0 \leq i < \ell} \in \mathcal{P}_{X_{v,a} Y_{v,a}}$ から $p = (B'_i)_{0 \leq i \leq \ell} \in \mathcal{P}_{XY}$ を作るために，p' において辺 b が交換される遷移に着目し，ある $j \in \{0, \ldots \ell\}$ に対して $b \in B'_j \oplus B'_{j+1}$ とする．このとき，$i \in \{0, \ldots, \ell\} \setminus \{j\}$ に対し，

$$B_i = \begin{cases} B'_i, & i < j \\ (B'_{i-1} \setminus \{a\}) \cup \{c\}, & i > j \end{cases}$$

とし，

$$B_j = \begin{cases} (B'_{j-1} \setminus \{a\}) \cup \{c\}, & (B'_{j-1} \setminus \{a\}) \cup \{c\} \\ & \cup A \text{ が森の場合} \\ (B'_{j-1} \setminus \{b\}) \cup \{c\}, & \text{otherwise} \end{cases}$$

とする．

以上のように多品種流を構成することで，定理 5 を得る．

5 まとめ・今後の課題

辺数が全域木よりも 1 つ少ない森上のマルコフ連鎖に対して，その混交時間が頂点数の多項式で抑えられることを示した．辺数が全域木よりも 1 つ多い連結全域部分グラフ上のマルコフ連鎖に対しても同様の結果が得られる．

今後の課題としては，任意の辺数の森上のマルコフ連鎖や連結全域部分グラフ上のマルコフ連鎖の混交時間が頂点数の多項式で抑えられるかどうか解析を行う必要がある．

参考文献

[1] R. Cordovil, M. L. Moreira, Bases-cobases graphs and polytopes of matroids, Combinatorica, 13 (1993), 157–165.

[2] J. Fehrenbach, L. Rüschendorf, Analysis of Markov chain algorithms on spanning trees, rooted forests, and connected subgraphs, Appl. Math., 32 (2005), 341–365.

[3] A. Sinclair, Improved bounds for mixing rates of Markov chains and multicommodity flow, Combinatorics, Probability and Computing, 1 (1992), 351–370.

[4] 伊里 正夫，藤重 悟，大山 達雄，グラフ・ネットワーク・マトロイド，産業図書，1986.