Vol.2018-AL-166 No.7
2018/1/29

シュタイナー三項系のサンプリングとマルコフ連鎖の高速混合

河本 和也[1,a)]　来嶋 秀治[1,2]

概要：本研究は，Steiner triple system(以下，STS) の高速な一様サンプリングが目的である．これまでに，マルコフ連鎖を用いて一様サンプリングを行うことを目的とした研究は存在するが，定数サイズ以上大きな頂点数に対して状態空間の既約性をもつマルコフ連鎖は知られていない．本発表では STS を拡張した状態空間上の一様分布を定常分布に持つマルコフ連鎖を構築し, 高速混合であることを示す. 拡張された状態空間の大きさは STS の数と比較して指数関数的に大きいが, 素朴な方法で構築できるマルコフ連鎖の状態空間と比べると指数的に小さい.

1. はじめに

本研究ではマルコフ連鎖を用いたシュタイナー三項系 (Steiner triple systems: STS) のサンプリングを行う問題を扱う．STS とはハイパーグラフの一種であり，少ない頂点数に対しても組み合わせ爆発によって多くの STS が存在することが知られている．Tonchev と Weishaar[5] は，頂点数 19 の STS が 11084874829 個存在することを証明した．これより頂点数の大きい STS の具体的な数は不明であるが，Wilson[6] や Linial と Luria[4] は頂点数 n の STS の数 $|\Xi_n|$ について，$\left(\frac{n}{e^2 3^{\frac{3}{2}}}\right)^{\frac{n^2}{6}} \leq |\Xi_n| \leq \left((1+o(1))\,\frac{n}{e^2}\right)^{\frac{n^2}{6}}$ であることを示している．

以上のように STS の具体的な数を知ることは困難であるため，マルコフ連鎖によるサンプリングを試みる先行研究が存在する．Cameron[1] は，STS と似た構造の和集合を状態空間とするマルコフ連鎖を提案し，Kaski ら [2] は，Switching(要素の入れ替え) による STS のみから成る状態空間上のマルコフ連鎖を提案している．しかしながら，いずれも大きな頂点数に対する既約性は未解決である．

本研究では，STS を拡張した状態空間上の一様分布を定常分布に持つマルコフ連鎖を構築し, 高速混合であることを示す. 本論文の構成は以下の通りである．2 章ではサンプリング対象である STS の定義を示す．3 章ではマルコフ連鎖の構築方法と，その定常分布を示す．4 章では 3 章で構築したマルコフ連鎖の混合時間を解析する．

1　九州大学 大学院システム情報科学府
2　JST さきがけ
a)　kazuya.kawamoto@inf.kyushu-u.ac.jp

2. 準備

2.1 Steiner triple system

STS とは，頂点集合 V の三つ組の集合 $T \subseteq \binom{V}{3}$ の事である．ただし V の各二つ組 (対) は，T の三つ組のちょうど一つに含まれる．頂点数 n の STS の集合 Ξ_n について，$n \equiv 1, 6 \pmod 6$ の時，かつその時に限り STS が存在すること，三つ組の数 $m = |T| = \frac{n(n-1)}{6}$ であること，各頂点の数 $s = \frac{n-1}{2}$ であることが知られている．また，$|\Xi_n|$ の上界・下界について，

$$\left(\frac{n}{e^2 3^{\frac{3}{2}}}\right)^{\frac{n^2}{6}} \leq |\Xi_n| \leq \left((1+o(1))\,\frac{n}{e^2}\right)^{\frac{n^2}{6}} \qquad (1)$$

である [4], [6].

2 つの STS に対し，各ブロックを要素とする頂点集合を U_1, U_2 とし，同じ対を有する 2 ブロックを要素とする辺集合を E として，2 部グラフ $G = (U_1, U_2; E)$ を定義する．この 2 部グラフについて，以下が成り立つ．

定理 1. 上述の $G = (U_1, U_2; E)$ は完全マッチングをもつ．

証明. ブロック $v_{1k} \in U_1$ とブロック $v_{2k} \in U_2$ が対を 2 つ以上共有すると仮定すると，各ブロックは 3 つの頂点から構成されるので，$v_{1k} = v_{2k}$ である．したがって，v_{1k}, v_{2k} が対を共有する際は必ず 1 つか 3 つの共通の対を有する．3 つ共通の頂点を有する組み合わせが存在する場合，その 2 ブロックをマッチングにおけるペアとみなせる．この時，各ブロックは 3 つの対を含んでいる事から残りの頂点の集合は，全ての頂点の次数が 3 である 2 部グラフとなる．Hall の結婚定理より，本グラフは完全マッチングを有する． □

Algorithm 1: Decide_next_state_MC_{STS}

```
1 if w.p. 1/2 then
2     i ← M からブロックをランダムに選択;
3     a ← m_i から要素をランダムに選択;
4     b ← V から要素をランダムに選択;
5     M' ← update(M, i, a, b) if M' ∈ Ξ'_n then
6         return M';
7 return M;
```

2.2 ラベル付き STS

ラベル付きSTSとは, ブロック間に順序があるとしたSTSの事である．すなわちラベル付きSTS $T = (t_1, t_2, ..., t_m) \in \binom{V}{3}^m$ は $\{t_1, ..., t_m\} \in \Xi_n$ を満たす．

定義 1. ラベル付き *STS* 全体の集合を $\Xi_n{}^* = \{T = (t_1, t_2, ..., t_m) | t_i \in \binom{V}{3}, \zeta(T) \in \Xi_n\}$ とする．ただし，$\zeta(T) = \bigcup_{i \in \{1,...,m\}} \{t_i\}$ とする．

任意のラベル付き STS $I = (i_1, ..., i_m) \in \Xi_n{}^*$ について，

$$\Xi_n{}^*(I) = \{T \in \Xi_n{}^* | \forall j \in \{1, ..., m\} | i_j \cap t_j| \geq 2\} \quad (2)$$

と定義する．

補題 2. 写像 $\zeta : \Xi_n{}^* \to \Xi_n$ は $\Xi_n{}^*(I)$ から Ξ_n への全射である．

証明. 定理 1 より，2 つの STS からなる 2 部グラフには必ず完全マッチングが存在する．これは，各ブロックに一つ存在する，マッチングに使用されていない頂点を変更することで，任意の STS に遷移できることを示している．したがって，あるラベル付き STS I について，各ブロックの頂点を高々一つ変更すれば任意の STS となるので，ζ は $\Xi_n{}^*(I)$ から Ξ_n への全射である． □

ただし，ζ は $\Xi_n{}^*(I)$ から Ξ_n への単射ではない．Ξ_n からの一様サンプリングを行うためには，$\Xi_n{}^*(I)$ からの一様サンプリングを行った後，Ξ_n から $\Xi_n{}^*(I)$ への単射である関数を用いて，その妥当性を判定する必要がある．そのような関数の一例を付録 A に示す．

3. $MC_{STS}(I)$

本章では，マルコフ連鎖 $MC_{STS}(I)$ を定め，その極限分布を議論する．

3.1 状態空間 $\Theta_n{}^*(I)$

$MC_{STS}(I)$ の状態空間は $\Theta_n{}^*(I) = \{T' \subseteq \binom{V}{3} | |T'| = m, \forall k \in \{1, ..., m\} | i_k \cap t'_k| \geq 2\}$ とする．明らかに $\Xi_n{}^*(I) \subset \Theta_n{}^*(I)$ である．

3.2 状態遷移

MC_{STS} の状態遷移は，現在の状態について 1 つの頂点を変更することで行う．状態 M について，i 番目のブロックに含まれる a を b に書き換えた状態を返す関数を

$$update(M, i, a, b) = (m_1, m_2, ..., (m_i \setminus a \cup b), ..., m_m) \quad (3)$$

とし，1 ステップの状態遷移は Algorithm1 に記載する．

3.3 定常分布

本節では，以下の定理を証明する．

定理 3. $\pi_M = 1/|\Theta_n{}^*(I)|$ で定義される確率分布 π は MC_{STS} の定常分布である．

補題 4. MC_{STS} が状態 M から $M'(\neq M)$ に遷移する確率 $p_{MM'}$ が *0* でないとき，$p_{MM'} = \frac{1}{6nm}$

証明. Algorithm1 の手順ごとに M から M' に遷移する確率を求める．

- 手順 1 M にとどまらない確率 1/2
- 手順 2 M' に遷移するブロックが選択される確率 $\frac{1}{m}$
- 手順 3 M' に遷移する要素が選択される確率 $\frac{1}{3}$
- 手順 4 M' に遷移する要素が選択される確率 $\frac{1}{n}$

したがって以上の積から，$p_{MM'} = \frac{1}{6nm}$． □

補題 5. 任意の M, M' について，詳細均衡条件 $p_{MM'} = p_{M'M}$ を満たす．

証明. M と M' の関係に従って，場合分けを行う．M と M' で 2 か所以上値が異なる場合, 一度の *update* 関数の適用によって遷移することはできないので，$p_{MM'} = p_{M'M} = 0$ である．また，M と M' で 1 か所のみ値が異なる場合，$M' = update(M, i, a, b)$ とすると $M = update(M', i, b, a)$ なので，4 より $p_{MM'} = p_{M'M} = \frac{1}{6nm}$．以上より，$p_{MM'} = p_{M'M}$． □

補題 6. MC_{STS} は既約である．

証明. 任意の状態 T' から初期状態 I に遷移可能であることを示す．任意の $k \in \{1, ..., m\}$ について $|i_k \cap t'_k| \geq 2$ より，各ブロック t'_k に i_k と異なる頂点 $i_k \setminus t'_k$ が高々一つ存在する．したがって，$\forall k \in \{1, ..., m\}$ について，$M_k = update(M_{k-1}, k, t'_k \setminus i_k, i_k \setminus t'_k)$ とすると，T' から I への経路 $\{T' = M_0, M_1, ..., M_m = I\}$ を得られる．各遷移において，M_k の j 番目のブロック $m_{kj} = i_j$ となるので，この遷移によって $\Theta_n{}^*(I)$ に属さない状態に遷移することはない．したがって，任意の状態 T' から I に遷移することが可能である． □

補題 7. MC_{STS} はエルゴード的である．

証明. Algorithm1 の手順 1 より，自己遷移確率 $p_{MM} \geq \frac{1}{2}$ であるため，MC_{STS} は非周期的である．補題 6 より，MC_{STS} は既約である．したがって，MC_{STS} はエルゴードである． □

Algorithm 2: γ_{XY}

1 $j \leftarrow 0$;
2 $M_0 \leftarrow X$;
3 **for** $k \in \{1, ..., n\}$ **do**
4 　**if** $|m_k \cap y_k| = 1$ **then**
5 　　$M_{j+1} \leftarrow update(M_j, i_k \setminus y_k, i_k \setminus m_k)$; $j++$;
6 　**if** $|m_k \cap y_k| = 2$ **then**
7 　　$M_{j+1} \leftarrow update(M_j, x_k \setminus y_k, y_k \setminus m_k)$; $j++$;

エルゴード的で詳細均衡条件を満たすことから ([3] 参照)，補題 5, 補題 7 より定理 3 が示される．

3.4 $\Theta_n{}^*(I)$ と Ξ_n の関係

$MC_{STS}(I)$ の状態空間の大きさ $|\Theta_n{}^*(I)|$ は，$|\Xi_n|$ の関係について言及しておく．

定理 8. $\frac{|\Xi_n|}{|\Theta_n{}^*(I)|} > \left(\frac{1}{3^{\frac{5}{2}}e^2}\right)^{n^2/6}$

証明. 式 (1) より，$|\Xi_n| \geq \left(\frac{n}{e^2 3^{\frac{3}{2}}}\right)^{\frac{n^2}{6}}$．また，$\Theta_n{}^*(I)$ の定義より，$T \in \Theta_n{}^*(I)$ の各ブロックについて，I と異なってよいのは高々一つの頂点である事，その値は高々 n 通りであることから，$|\Theta_n{}^*(I)| < 3^m n^m < (3n)^{n^2/6}$．以上より，$\frac{|\Xi_n|}{|\Theta_n{}^*(I)|} > \left(\frac{1}{3^{\frac{5}{2}}e^2}\right)^{n^2/6}$ □

4. $MC_{STS}(I)$ の混合時間

本章では，$MC_{STS}(I)$ の混合時間について以下の定理を示す．

定理 9. $MC_{STS}(I)$ の混合時間 $\tau(\delta) = O(n^6(n^2 \log n + \log \delta^{-1}))$

まず，混合時間の見積もりのために必要な補題を示す．

補題 10. $X = (x_1, x_2, ..., x_m), Y = (y_1, y_2, ..., y_m) \in \Theta_n{}^*(I)$ について，$\forall k \in \{1, ..., n\}, |x_k \cap y_k| \geq 1$

証明. $\Theta_n{}^*(I)$ の定義より，$\forall k \in \{1, ..., n\}|i_k \cap x_k| \geq 2$. $\forall k \in \{1, ..., n\}|x_k \cap y_k| = 0$ と仮定すると，$\forall k \in \{1, ..., n\}|i_k \cap y_k| \leq 1$ となり，$\Theta_n{}^*(I)$ の定義と矛盾する．したがって，補題が成り立つ． □

$X \in \Theta_n{}^*(I)$ から $Y \in \Theta_n{}^*(I)$ への正準経路 $\gamma_{XY} : X = M_0, M_1, ..., M_k = Y$ を Algorithm 2 のように構成する．

補題 11. γ_{XY} に従ったとき, 遷移する全ての状態は $\Theta_n{}^*(I)$ に属する．

証明. まず，4 行目の遷移については，除外する頂点，追加する頂点共に i_k に属するので，$m_k \cap i_i$ に属する頂点が減少することはない．また 5 行目の条件より，x_k と y_k は異なる頂点を一つ有し，6 行目の遷移によって $m_k = y_k$ となることがわかる．$|i_k \cap y_k| \geq 2$ より，$|i_k \cap m_k| \geq 2$．したがって，補題が成り立つ． □

補題 12. 任意の $X = \{x_1, x_2, ..., x_m\}, Y = \{y_1, y_2, ..., y_m\} \in \Theta_n{}^*(I)$ について，γ_{XY} は X から Y への経路である．

証明. $|m_k \cap y_k| = 1$ より，$|i_k \cap y_k \setminus m_k| = 1$. したがって，$i_k \setminus m_k \in y_k$．また，$i_k \setminus y_k \notin y_k$ なので，$|m_k \cap y_k| = 1$ である時，4 行目の遷移によって $|m_k \cap y_k| = 2$ となる．さらに，$y_k \setminus x_k \in y_k$，$x_k \setminus y_k \notin y_k$ より，$|m_k \cap y_k| = 2$ である時，6 行目の遷移によって $|m_k \cap y_k| = 3$ となる．したがって，$|m_k \cap y_k|$ がどのような値でも，$m_k = y_k$ となる． □

補題 11 と補題 12 から，γ_{XY} は X から Y への正準経路となっている．

γ_{XY} の集合 Γ の混雑度 $\rho(\Gamma)$ を計算する．正準経路中に MM' を含むような X, Y の集合を $cp(M, M') = \{(X, Y)|(M, M') \in \gamma_{XY}\}$ として，写像 $\eta_{MM'} : cp(M, M') \to \Xi_n$ を，

$$\eta_{MM'}(X, Y) = \bigcup_{k \in \{1, ..., m\}} (x_k \cap y_k) \cup ((x_k \oplus y_k) \setminus m_k) \quad (4)$$

と定義する．

補題 13. $\eta_{MM'}$ は単射である．

証明. $m_k, m'_k, \eta_{MM'}(X, Y)k$ から，x_k, y_k を復元できることを示す．簡単のため，$\eta_k = \eta_{MM'}(X, Y)k$ とする．$\eta_{MM'}$ の定義より，$x_k \cap y_k = \eta_k \cap m_k$，$x_k \oplus y_k = \eta_k \oplus m_k$．したがって，$\eta_k \oplus m_k$ の各頂点が x_k, y_k のどちらに属するかを判別できればよい．

まず，Algorithm2 が操作するブロックの順番は決まっているので，M, M' 間で操作されていないブロックに関しては，操作済みであるかの判定が可能である．M, M' 間で操作されているブロックについては，場合分けして考える．

$|m_k \cap \eta_k| = 2$ である場合は，M, M' 間で操作されている頂点以外，そのブロックで操作されるものはない．$|m_k \cap \eta_k| = 1$ である場合は，2 つの頂点を操作する必要がある．このブロックは，Algorithm 2 における 4 行目と 6 行目両方で遷移を行うことになるが，4 行目では除外・追加する頂点共に i_k に属するのに対し (図 1)，6 行目の操作では属していない (図 2)．この性質を用いて，$\eta_k \oplus m_k$ に属しているが，M, M' 間で操作対象となっていない頂点が操作済みであるかどうかの判定が可能である．

以上の議論より全ての頂点が操作済みであるか否かを判定できるので，m_k に属し操作済みでない頂点と，η_k に属し操作済みである頂点は x_k に，その他の頂点は y_k に属することがわかる．したがって，$m_k, m'_k, \eta_{MM'}(X, Y)k$ から，x_k, y_k を復元できる． □

補題 14. $\rho(\Gamma) = O(n^3)$

Vol.2018-AL-166 No.7
2018/1/29

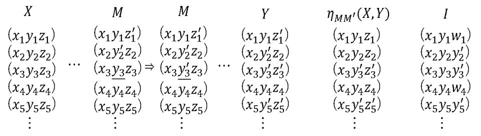

図 1: x, y の復元 1

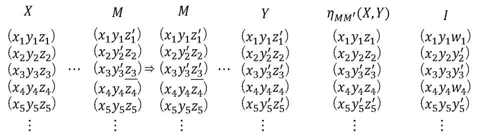

図 2: x, y の復元 2

証明.

$$
\begin{aligned}
\rho(\Gamma) &= \max_{(M,M')\in E(\Omega_{mn})} \frac{1}{\pi_M p_{MM'}} \sum_{(X,Y)\in cp(M,M')} \pi_X \pi_Y \\
&= \max_{(M,M')\in E(\Omega_{mn})} \frac{1}{\frac{1}{|\Theta_n|}\frac{1}{6nm}} \sum_{(X,Y)\in cp(M,M')} \frac{1}{|\Theta_n|} \pi_{\eta_{MM'}(X,Y)} \\
&\quad \text{補題 13 より,} \qquad (5) \\
&= \max_{(M,M')\in E(\Omega_{mn})} \frac{1}{\frac{1}{|\Theta_n|}\frac{1}{6nm}} \frac{1}{|\Theta_n|} \\
&= \frac{n^2(n-1)}{2} \\
&= O(n^3)
\end{aligned}
$$

□

Jerrum と Sinclair は，混雑度と混合時間の関係について，以下を示している ([3] 参照).

定理 15. Ω を状態空間とするマルコフ連鎖の混合時間 $\tau(\delta) = O(\rho(\Gamma)^2(\log|\Omega| + \log\delta^{-1}))$.

以上より，$MC_{STS}(i)$ の混合時間の上界が示せる.

定理 9 の証明.

$$
\begin{aligned}
\tau(\delta) &= O(\rho(\Gamma)^2 \log|\Omega| + \log\delta^{-1}) \\
&= O(n^6(\frac{n^2}{6}\log n + \log\delta^{-1})) \\
&= O(n^6(n^2\log n + \log\delta^{-1}))
\end{aligned}
$$

□

5. おわりに

本研究では，STS の一様サンプリングを行うためのマルコフ連鎖を示し，大きな頂点数に対しても既約性を保証し，高速混合であることを示した．今後の課題は，STS の数に対して指数関数的に大きい状態空間の縮小と，混合時間の短縮である．

謝辞

本研究は JST さきがけ JPMJPR16E4 の支援を受けている．

Algorithm 3: $get\_canonical\_state_I(T)$

```
mat ← find_matching(ζ(I), T)
for k ∈ 1, 2, ..., m do
    t*_k ← find_counterpart(mat, i_k);
return {t*_1, t*_2, ..., t*_m};
```

付　　録

付録 A　Ξ_n から $\Xi_n{}^*(I)$ への単射

Ξ_n から $\Xi_n{}^*(I)$ への単射である関数の一例を Algorithm 3 に示す．ここで，*find_matching* は 2 つの STS からなる 2 部グラフにおける，完全マッチングを決定的に一つ探索する関数であり，$find\_counterpart$ はあるマッチングにおけるある頂点の片割れを探索する関数である．

補題 16. $T^* = get\_canonical\_state_I(T)$ とすると，$T = \zeta(T^*)$

証明.

$$
\begin{aligned}
(\text{右辺}) = \zeta(T^*) &= \bigcup_{i\in\{1,\dots,m\}} t_k^* \\
&= \bigcup_{i\in\{1,\dots,m\}} find\_counterpart(mat, i_k) \\
&= \bigcup_{i\in\{1,\dots,m\}} t_k = T = (\text{左辺})
\end{aligned}
$$

□

定理 17. $get\_canonical\_state_I(T)$ は Ξ_n から $\Xi_n{}^*(I)$ への単射である．

証明. 任意の $T_1, T_2 \in \Xi_n(T_1 \neq T_2)$ について，$get\_canonical\_state_I(T_1) = get\_canonical\_state_I(T_2) = T^*$ と仮定する．この時，補題 16 より，$T_1 = T_2 = \zeta(T^*)$. これは仮定と矛盾するので，$get\_canonical\_state_I$ は単射である． □

参考文献

[1] P. J. Cameron, A Markov chain for Steiner triple systems, Working document, 2002.
[2] P. Kaski, V. Mäkinen, P. R. J. Östergard, The cycle switching graph of the Steiner triple systems of order 19 is connected, Graphs and Combinatorics, 27:539, 2011.
[3] D. A. Levin, Y. Peres, E. L. Wilmer, Markov Chains and Mixing Times, American Mathematical Soc, 2009.
[4] N. Linial and Z. Luria, Upper bounds on the number of Steiner triple systems and 1-factorizations, Random Struct. Alg. 43:399–406, 2013.
[5] V. D. Tonchev, and R. S. Weishaar, Steiner triple systems of order 15 and their codes. J. Stat. Plan. Inference 58, 207–216, 1997.
[6] M. R. Wilson, Nonisomorphic Steiner triple systems, Math. Z. 135, 303–313, 1974.