2004－AL－94　(10)
2004／3／19

離散化 Dirichlet 分布に従うパーフェクトサンプリング

松井 知己 *　来嶋 秀治*

概要

本稿では離散化 Dirichlet 分布に厳密に従うサンプリングの手法を提案する。提案するアルゴリズムはマルコフ連鎖を用いたサンプリング法で、単調 CFTP (Coupling From The Past) アルゴリズムに基づく Las Vegas 型の乱択アルゴリズム（randomized algorithm）である。我々は離散化 Dirichlet 分布を唯一の極限分布に持つ新しいマルコフ連鎖を提案する。このマルコフ連鎖の各推移では離散化ベータ分布に従う確率変数を生成し、アルゴリズムは平均 $O(n^3 \ln \Delta)$ 回の推移で終了して、Dirichlet 分布に厳密に従う確率変数を返す。ただし、n は変数の次元（パラメータの個数）であり、$1/\Delta$ は離散化のグリッド幅の大きさである。従って、計算時間はパラメータの数値の大きさに依存しない。

Perfect Sampler for Discretized Dirichlet Distribution

Tomomi MATSUI and Shuji KIJIMA

Abstract

In this paper, we propose a perfect (exact) sampling algorithm according to a discretized Dirichlet distribution. The Dirichlet distribution appears as prior and posterior distribution for the multinomial distribution in many statistical methods in bioinformatics. Our algorithm is a monotone coupling from the past algorithm, which is a Las Vegas type randomized algorithm. We propose a new Markov chain whose limit distribution is a discretized Dirichlet distribution. Our algorithm simulates transitions of the chain $O(n^3 \ln \Delta)$ times where n is the dimension (the number of parameters) and $1/\Delta$ is the grid size for discretization. Thus the obtained bound does not depend on the magnitudes of parameters. In each transition, we need to sample a random variable according to a discretized beta distribution (2-dimensional Dirichlet distribution).

1　はじめに

本稿では離散化 Dirichlet 分布に厳密に従うサンプリングの手法を提案する。提案するアルゴリズムはマルコフ連鎖を用いたサンプリング法で、単調 CFTP (Coupling From The Past) アルゴリズムに基づく Las Vegas 型の乱択アルゴリズム（randomized algorithm）である。

Dirichlet 分布は生物情報学の多くの統計的手法における多項分布の事前、事後分布としてしばしば現れる。生物情報学において、観測データから（共通の）疾患を引き起こす遺伝子を見つけるための強力な道具として統計的手法は広く用いられている。これらの手法として EM アルゴリズム、マルコフ連鎖モンテカルロ法、Gibbs サンプラーがあげられる。Dirichlet 分布は多項分布におけるパラメータの共役事前分布な

*東京大学 大学院情報理工学系研究科 数理情報学専攻
Department of Mathematical Informatics, Graduate School of Information Science and Technology, University of Tokyo,
http://www.simplex.t.u-tokyo.ac.jp/~tomomi/
kijima@misojiro.t.u-tokyo.ac.jp

ので、これらの手法の多項分布の事前あるいは事後確率としてしばしば現れる [16]。たとえば、Niu, Qin, Xu, and Liu は各項目に対するディプロタイプの配置を確率的に定めるベイジアンハプロタイプ推定法を提案している。他の例として、 Pritchard, Stephens, and Donnely の母集団構造推定アルゴリズムが挙げられる [13]。このアルゴリズムでは MCMC 法を用いている。これらの例において、Dirichlet 分布は様々な次元で、様々なパラメータをとって現れる。したがって、任意の次元とパラメータを持つ Dirichlet 分布からサンプリングを行える効率的アルゴリズムが望まれる。

（連続）Dirichlet 分布からのサンプリングのひとつの方法としては棄却サンプリングがある（[6] 参照）。しかしパラメータが小さいと、$n = 2$ の場合（ベータ分布）でさえ棄却の確率はすぐに大きくなってしまいパラメータ値が小さいとき、効率的でない。他の手法として、領域を離散化し、Metropolis-Hastings アルゴリズムを用いる手法がある。近年、Matsui, Motoki and Kamatani が離散化 Dirichlet 分布に従うサンプリングのためのマルコフ連鎖を提案している [11]。そのマルコフ連鎖の mixing time は $(1/2)n(n-1)(1+\ln(\Delta-n))\log\varepsilon$ で押さえられる。

CFTP アルゴリズムは Propp and Wilson によって提案され、画期的アルゴリズムとして注目を浴びている [14]。このアルゴリズムでは、マルコフ連鎖のシミュレーションを工夫することで定常分布に厳密に従うサンプリング（Perfect sampling）を可能とする。Perfect sampling を行う利点は、定常分布に厳密に従うサンプリングを行うことで、誤差パラメータを考慮する必要が無くなる点である。特に精度の高いサンプリングを要する時、Perfect sampling は近似サンプリングよりも速いアルゴリズムとなる。しかし CFTP アルゴリズムは、そのままではマルコフ連鎖の全状態数に比例する計算量を必要とするため、状態数の多い対象に対して効率的ではない。対象とするマルコフ連鎖に「単調性」がある時、効率的な CFTP アルゴリズムの設計が可能となる。これを単調 CFTP アルゴリズムと呼ぶ。一般に単調なマルコフ連鎖の設計は困難で、これまで実際に単調 CFTP アルゴリズムの設計された例は少ない。

本稿では離散化 Dirichlet 分布を唯一の極限分布に持つ新しいマルコフ連鎖を提案する。提案するマルコフ連鎖は単調である。このマルコフ連鎖の各推移では離散化ベータ分布に従う確率変数を生成し、アルゴリズムは平均 $\mathrm{O}(n^3 \ln\Delta)$ 回の推移で終了して、Dirichlet 分布に厳密に従う確率変数を返す。ただし、n は変数の次元（パラメータの個数）であり、$1/\Delta$ は離散化のグリッド幅の大きさである。従って、計算時間はパラメータの数値の大きさに依存しない。提案するマルコフ連鎖の各推移の計算量は [11] と同じで、アルゴリズム中で必要な推移の期待回数は $\mathrm{O}(n^3 \ln\Delta)$ である。従って、n が $\mathrm{O}(\ln\varepsilon)$ より小さいとき、今回提案する手法は近似法よりも速く、しかも得られる確率変数ベクトルは定常分布に厳密に従う。

2　サンプリングアルゴリズム

非負実数パラメータ $u_1, \ldots, u_n$ を持つ Dirichlet 分布は、確率変数ベクトル $P = (P_1, P_2, \ldots, P_n)$ に対する確率分布で、密度関数は定義域 $\{(p_1, p_2, \ldots, p_n) \in \mathbb{R}^n \mid \sum_{i=1}^n p_i,\ p_i > 0\ (\forall p_i)\}$ に対して

$$\frac{\Gamma(\sum_{i=1}^n u_i)}{\prod_{i=1}^n \Gamma(u_i)} \prod_{i=1}^n p_i^{u_i-1}$$

で表される。ただし、$\Gamma(u)$ はガンマ関数である。本論文では $n \geq 2$ を仮定する。

任意の整数 $\Delta \geq n$ に対して、定義域を格子幅 $1/\Delta$ で離散化し、整数ベクトルの離散的集合 Ω を

$$\Omega \stackrel{\text{def.}}{=} \{(x_1, x_2, \ldots, x_n) \in \mathbb{Z}_{++}^n \mid x_i > 0\ (\forall i),\ x_1 + \cdots + x_n = \Delta\}$$

で定義する。非負実数パラメータ $u_1, \ldots, u_n$ を持つ離散化 Dirichlet 確率変数は確率変数ベクトル $X = (X_1, \ldots, X_n) \in \Omega$ で確率分布

$$\Pr[X = (x_1, \ldots, x_n)] \stackrel{\text{def.}}{=} C_\Delta \prod_{i=1}^n (x_i/\Delta)^{u_i-1}$$

を持つ。ただし、C_Δ は分配関数（規格化定数）で $(C_\Delta)^{-1} \stackrel{\text{def.}}{=} \sum_{\boldsymbol{x}\in\Omega} \prod_{i=1}^{n} (x_i/\Delta)^{u_i-1}$ と定義される。

任意の整数 $b \geq 2$ に対して、2 次元整数ベクトルの集合 $\Omega(b) \stackrel{\text{def.}}{=} \{(Y_1, Y_2) \in \mathbb{Z}^2 \mid Y_1, Y_2 > 0,\ Y_1 + Y_2 = b\}$ を導入し、非負実数パラメータ u_i, u_j を持つ分布関数 $f_b(Y_1, Y_2 \mid u_i, u_j) : \Omega(b) \to [0, 1]$ を

$$f_b(Y_1, Y_2 \mid u_i, u_j) \stackrel{\text{def.}}{=} C(u_i, u_j, b) Y_1^{u_i-1} Y_2^{u_j-1}$$

と定義する。ただし、定数 $(C(u_i, u_j, b))^{-1} \stackrel{\text{def.}}{=} \sum_{(Y_1,Y_2)\in\Omega(b)} Y_1^{u_i-1} Y_2^{u_j-1}$ は分配関数である。またベクトル $(g_b(0|u_i, u_j), g_b(1|u_i, u_j), \ldots, g_b(b-1|u_i, u_j))$ を導入し、

$$g_b(k|u_i, u_j) \stackrel{\text{def.}}{=} \begin{cases} 0 & (k = 0) \\ \sum_{l=1}^{k} C(u_i, u_j, b) l^{u_i-1} (b-l)^{u_j-1} & (k \in \{1, 2, \ldots, b-1\}) \end{cases}$$

と定義する。明らかに $0 = g_b(0|u_i, u_j) < g_b(1|u_i, u_j) < \cdots < g_b(b-1|u_i, u_j) = 1$ が成り立つ。

状態空間 Ω を持つマルコフ連鎖 $\mathcal{M}_{\mathrm{D}}$ について述べる。現在の状態を $X \in \Omega$ とする。このとき、推移 $X \mapsto X'$ は次のように実行される。まず、実数乱数 $\lambda \in [1, n)$ を生成し、$i := \lfloor \lambda \rfloor$, $b := X_i + X_{i+1}$ とする。次に、$k \in \{1, 2, \ldots, b-1\}$ を $g_b(k-1|u_i, u_{i+1}) \leq (\lambda - \lfloor \lambda \rfloor) < g_b(k|u_i, u_{i+1})$ を満たす唯一の値とする。最後に、

$$X'_j := \begin{cases} k & (j = i), \\ b - k & (j = i+1), \\ X_j & (\text{それ以外}), \end{cases}$$

とする。

提案したマルコフ連鎖に対する update function $\phi : \Omega \times [1, n) \to \Omega$ を $\phi(X, \lambda) \stackrel{\text{def.}}{=} X'$ とする。このマルコフ連鎖は明らかに既約で非周期的である。また detailed balance equations が成り立つことから、マルコフ連鎖 $\mathcal{M}_{\mathrm{D}}$ の定常分布は離散化 Dirichlet 分布となる。

次に特別な状態として $X_{\mathrm{U}}, X_{\mathrm{L}} \in \Omega$ を

$$X_{\mathrm{U}} \stackrel{\text{def.}}{=} (\Delta - n + 1, 1, 1, \ldots, 1), \quad X_{\mathrm{L}} \stackrel{\text{def.}}{=} (1, 1, \ldots, 1, \Delta - n + 1).$$

で定義する。

これらを用いて、離散化 Dirichlet 分布に対するサンプリングアルゴリズムを次のように定める。

アルゴリズム 1

Step 1. 時間に関する初期値を $T := -1$ とし、過去に遡る。空列 $\boldsymbol{\lambda}$ を用意する。

Step 2. 一様実数乱数 $\lambda[T], \lambda[T+1], \ldots, \lambda[\lceil T/2 \rceil - 1] \in [1, n)$ を生成し、$\boldsymbol{\lambda} := (\lambda[T], \lambda[T+1], \ldots, \lambda[-1])$ とする。

Step 3. 時刻 T における 2 本のマルコフ連鎖の状態をそれぞれ X_{U} と X_{L} し、共通の数列 $\boldsymbol{\lambda}$ を用いて update function ϕ に従い、マルコフ連鎖を時刻 T から時刻 0 に至るまで推移させる。

Step 4. [Coalescence check]

(a) もし $\exists Y \in \Xi,\ Y = \Phi_T^0(X_{\mathrm{U}}, \boldsymbol{\lambda}) = \Phi_T^0(X_{\mathrm{L}}, \boldsymbol{\lambda})$, ならば値 Y を返し、停止する。

(b) もしそうでなければ、時刻を $T := 2T$ として Step 2 に戻る。

以下は本稿の主要な結果である。

定理 1 アルゴリズム 1 は確率 1 で（有限時間で）停止して、状態を 1 つ出力する。その状態は Ω 上の離散化 Dirichlet 分布に厳密に従う確率変数の実現値である。

条件 1 Dirichlet パラメータは非増加順に並ぶ。すなわち $u_1 \geq u_2 \geq \cdots \geq u_n$ が成り立つ。

定理 2 条件 1 の下、アルゴリズム 1 の計算時間の期待値は $\mathrm{O}(n^3 \ln \Delta)$ である。ただし、n は次元（Dirichlet パラメータの個数）であり、$1/\Delta$ は離散化の格子幅を表す。

3　マルコフ連鎖の単調性

本章ではでは Ω 上に半順序を導入し、マルコフ連鎖 $\mathcal{M}_{\mathrm{D}}$ が単調であることを示す。任意のベクトル $X \in \Omega$ に対して、**累積和ベクトル** $\mathrm{c}_X \in \mathbb{Z}_+^{n+1}$

$$\mathrm{c}_X(i) \stackrel{\mathrm{def.}}{=} \begin{cases} 0 & (i=0), \\ X_1 + X_2 + \cdots + X_i & (i \in \{1,2,\ldots,n\}), \end{cases}$$

で定義する。ただし、$\mathrm{c}_X = (\mathrm{c}_X(0), \mathrm{c}_X(1), \ldots, \mathrm{c}_X(n))$ とする。明らかに Ω と $\{\mathrm{c}_X \mid X \in \Omega\}$ の間には一対一対応が存在する。任意の状態対 $X, Y \in \Omega$ に対して、$X \succeq Y$ の必要十分条件を $\mathrm{c}_X \geq \mathrm{c}_Y$ とする。明らかに "$\succeq$" は Ω 上の半順序である。また、$\forall X \in \Omega$, $X_{\mathrm{U}} \succeq X \succeq X_{\mathrm{L}}$ も簡単に分かる。

いま $X, Y \in \Omega$ に対して、ある k が存在して

$$\mathrm{c}_X(i) - \mathrm{c}_Y(i) = \begin{cases} 1 & (i=j), \\ 0 & (\text{otherwise}). \end{cases}$$

が成り立つとき、X が（k で）Y を**被覆**（*cover*）すると言い、$X \cdot\!\succ Y$（または $X \cdot\!\succ_k Y$）であらわす。

次の補題は単調性を証明する鍵となる。

補題 3 もし 2 つの相異なる状態 $X, Y \in \Omega$ が $X \cdot\!\succ_k Y$ ならば、$\forall \lambda \in [1,n)$, $\phi(X,\lambda) \succeq \phi(Y,\lambda)$ が成り立つ。

証明: 簡単のため $X' = \phi(X,\lambda)$ および $Y' = \phi(Y,\lambda)$ で表す。まず、任意の添え字 $i \neq \lfloor\lambda\rfloor$ に対しては $\mathrm{c}_X(i) = \mathrm{c}_{X'}(i)$ および $\mathrm{c}_Y(i) = \mathrm{c}_{Y'}(i)$ が成り立つ。定理の仮定より $X \succeq Y$ なので $\mathrm{c}_{X'}(i) - \mathrm{c}_{Y'}(i) = \mathrm{c}_X(i) - \mathrm{c}_Y(i) \geq 0$ である。以下では $\mathrm{c}_{X'}(\lfloor\lambda\rfloor) \geq \mathrm{c}_{Y'}(\lfloor\lambda\rfloor)$ を示す。

マルコフ連鎖 $\mathcal{M}_{\mathrm{D}}$ の定義から、$X'_{\lfloor\lambda\rfloor}$ の値は

$$g_{b'}(k'-1|u_{\lfloor\lambda\rfloor}, u_{\lfloor\lambda\rfloor+1}) \leq (\lambda - \lfloor\lambda\rfloor) < g_{b'}(k'|u_{\lfloor\lambda\rfloor}, u_{\lfloor\lambda\rfloor+1})$$

を満たす唯一の k' である。ただし、$b' \stackrel{\mathrm{def.}}{=} X_{\lfloor\lambda\rfloor} + X_{\lfloor\lambda\rfloor+1}$ である。同様に、$Y'_{\lfloor\lambda\rfloor}$ の値は

$$g_{b''}(k''-1|u_{\lfloor\lambda\rfloor}, u_{\lfloor\lambda\rfloor+1}) \leq (\lambda - \lfloor\lambda\rfloor) < g_{b''}(k''|u_{\lfloor\lambda\rfloor}, u_{\lfloor\lambda\rfloor+1})$$

を満たす唯一の k'' である。ただし、$b'' \stackrel{\mathrm{def.}}{=} Y_{\lfloor\lambda\rfloor} + Y_{\lfloor\lambda\rfloor+1}$ である。証明では次の 3 つの場合に分けて考える。

Case 1: $\lfloor\lambda\rfloor \neq j-1$ かつ $\lfloor\lambda\rfloor \neq j+1$ の場合、$b' = b''$ なので $X'_{\lfloor\lambda\rfloor} = k' = k'' = Y'_{\lfloor\lambda\rfloor}$ を得る。

Case 2: $\lfloor\lambda\rfloor = j-1$ の場合を考える。いま、$X \cdot\!\succ_j Y$ より $b' = b'' + 1$ である。累積和ベクトルの定義から

$$\begin{aligned} \mathrm{c}_{X'}(j-1) - \mathrm{c}_{Y'}(j-1) &= \mathrm{c}_{X'}(j-2) + X'_{j-1} - \mathrm{c}_{Y'}(j-2) - Y'_{j-1} \\ &= \mathrm{c}_X(j-2) + X'_{j-1} - \mathrm{c}_Y(j-2) - Y'_{j-1} = X'_{j-1} - Y'_{j-1} \end{aligned}$$

なので、$X'_{j-1} \geq Y'_{j-1}$ を示せば十分である。

いま、この後に記す補題 4 から、次の不等式が得られる。

$$\begin{aligned} 0 &= g_{b''+1}(0|u_{j-1}, u_j) = g_{b''}(0|u_{j-1}, u_j) \leq g_{b''+1}(1|u_{j-1}, u_j) \leq g_{b''}(1|u_{j-1}, u_j) \leq \cdots \\ &\leq g_{b''+1}(k-1|u_{j-1}, u_j) \leq g_{b''}(k-1|u_{j-1}, u_j) \leq g_{b''+1}(k|u_{j-1}, u_j) \leq \cdots \\ &\leq g_{b''+1}(b''-1|u_{j-1}, u_j) \leq g_{b''}(b''-1|u_{j-1}, u_j) = g_{b''+1}(b''|u_{j-1}, u_j) = 1, \end{aligned}$$

この不等式を *alternating inequalities* と呼ぶことにする。たとえば不等式

$$g_{b''+1}(k-1|u_{j-1}, u_j) \leq (\lambda - \lfloor\lambda\rfloor) < g_{b''}(k-1|u_{j-1}, u_j) \leq g_{b''+1}(k|u_{j-1}, u_j)$$

が成り立つ時、$X'_{\lfloor\lambda\rfloor} = k > k-1 = Y'_{\lfloor\lambda\rfloor}$ である。あるいは、不等式

$$g_{b''+1}(k-1|u_{j-1},u_j) \le g_{b''}(k-1|u_{j-1},u_j) \le (\lambda - \lfloor\lambda\rfloor) < g_{b''+1}(k|u_{j-1},u_j)$$

が成り立てば $X'_{\lfloor\lambda\rfloor} = k = Y'_{\lfloor\lambda\rfloor}$ である。すなわち alternating inequalitis から、任意の λ について coupling

$$\begin{pmatrix} X'_{j-1} \\ Y'_{j-1} \end{pmatrix} \in \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \begin{pmatrix} 3 \\ 2 \end{pmatrix}, \ldots, \begin{pmatrix} b''-1 \\ b''-1 \end{pmatrix}, \begin{pmatrix} b'' \\ b''-1 \end{pmatrix} \right\}$$

を構成することができる。以上のことから $X'_{j-1} \ge Y'_{j-1}$ が得られた。

Case 3: $\lfloor\lambda\rfloor = j+1$ の場合について考える。いま、$X \cdot\succ_j Y$ より $b'+1 = b''$ である。累積和ベクトルの定義から

$$\begin{aligned} \mathrm{c}_{X'}(j+1) - \mathrm{c}_{Y'}(j+1) &= \mathrm{c}_{X'}(j) + X'_{j+1} - \mathrm{c}_{Y'}(j) - Y'_{j+1} \\ &= \mathrm{c}_X(j) + X'_{j+1} - \mathrm{c}_Y(j) - Y'_{j+1} = 1 + X'_{j+1} - Y'_{j+1}. \end{aligned}$$

なので、$1 + X'_{j+1} \ge Y'_{j+1}$ を示せば十分である。

補題 4 から、Case 2 と同様に次の alternating inequalities が得られる。

$$\begin{aligned} 0 &= g_{b'+1}(0|u_{j+1},u_{j+2}) = g_{b'}(0|u_{j+1},u_{j+2}) \\ &\le g_{b'+1}(1|u_{j+1},u_{j+2}) \le g_{b'}(1|u_{j+1},u_{j+2}) \le \cdots \\ &\le g_{b'+1}(k-1|u_{j+1},u_{j+2}) \le g_{b'}(k-1|u_{j+1},u_{j+2}) \le g_{b'+1}(k|u_{j+1},u_{j+2}) \le \cdots \\ &\le g_{b'+1}(b'-1|u_{j+1},u_{j+2}) \le g_{b'}(b'-1|u_{j+1},u_{j+2}) = g_{b'+1}(b'|u_{j+1},u_{j+2}) = 1. \end{aligned}$$

したがって任意の λ について coupling (X', Y') は

$$\begin{pmatrix} X'_{j+1} \\ Y'_{j+1} \end{pmatrix} \in \left\{ \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ 2 \end{pmatrix}, \begin{pmatrix} 2 \\ 3 \end{pmatrix}, \ldots, \begin{pmatrix} b'-1 \\ b'-1 \end{pmatrix}, \begin{pmatrix} b'-1 \\ b' \end{pmatrix} \right\}$$

であることがわかる。以上のことから $1 + X'_{j+1} \ge Y'_{j+1}$ が得られた。 □

補題 4 ([10]) $\forall b \in \{2,3,\ldots\},\ \forall u_i, \forall u_j \ge 0,\ \forall k \in \{1,2,\ldots,b\}$,

$$g_{b+1}(k-1|u_i,u_j) \le g_b(k-1|u_i,u_j) \le g_{b+1}(k|u_i,u_j).$$

□

補題 5 update function ϕ で定義されたマルコフ連鎖 $\mathcal{M}_{\mathrm{D}}$ は "$\succeq$" に関して単調である。すなわち $\forall\lambda \in [1,n)$, $\forall X, \forall Y \in \Omega,\ X \succeq Y \Rightarrow \phi(X,\lambda) \succeq \phi(Y,\lambda)$ が成り立つ。

証明: いま、適当な長さの状態列 $Z_1, Z_2, \ldots, Z_r$ が存在して、$X = Z_1 \cdot\succ Z_2 \cdot\succ \cdots \cdot\succ Z_r = Y$ が成り立つ。したがって、補題 3 を繰り返し適用することで $\phi(X,\lambda) = \phi(Z_1,\lambda) \succeq \phi(Z_2,\lambda) \succeq \cdots \succeq \phi(Z_r,\lambda) = \phi(Y,\lambda)$ が得られる。 □

4 アルゴリズムの計算時間

本章ではアルゴリズムの計算時間について議論する。以下、coalescence time $T_* \in \mathbb{Z}_{++}$ を算定することで定理 2 を示す。ただし、coalescence time は $T_* \stackrel{\text{def.}}{=} \min\{t > 0 \mid \exists y \in \Omega,\ \forall x \in \Omega,\ y = \Phi^0_{-t}(x,\mathbf{\Lambda})\}$ と定義される確率変数である。

補題 6 条件 1 の下で、マルコフ連鎖 $\mathcal{M}_{\mathrm{D}}$ の coalescence time は $\mathrm{E}[T_*] = \mathrm{O}(n^3 \ln\Delta)$ を満たす。

証明: 単純無向グラフ $G=(\Omega,\mathcal{E})$ は頂点集合 Ω と次に定義する枝の集合 $\mathcal{E}$ を持つ。任意の頂点対 $\{X,Y\}$ が $\mathcal{E}$ に属する必要十分条件は $(1/2)\sum_{i=1}^{n}|X_i-Y_i|=1$ とする。明らかにグラフ G は連結である。各枝 $e=\{X,Y\}\in\mathcal{E}$ に対して、唯一の添え字対 $j_1,j_2\in\{1,\ldots,n\}$ が存在して

$$|X[1,j]-Y[1,j]|=\begin{cases}1 & (j=j_1,j_2),\\ 0 & (\text{otherwise}).\end{cases}$$

を満たす。この添え字の対を枝 e の**支持対**（*suporting pair*）と呼ぶ。枝 e の支持対 $\{j_1,j_2\}$ に対して $j^*=\max\{j_1,j_2\}\geq 2$ として、枝 e の長さ $l(e)$ を $l(e)\stackrel{\text{def.}}{=}(1/(n-1))\sum_{i=1}^{j^*-1}(n-i)$ で定義する。ここで $1\leq\min_{e\in\mathcal{E}}l(e)\leq\max_{e\in\mathcal{E}}l(e)\leq n/2$ に注意が必要である。任意の状態対 $X,Y\in\Omega$ に対して、距離 $d(X,Y)$ はグラフ G 上での X と Y の最短経路と定義する。任意の $(X,Y)\in\Omega^2$ に対して $d(X,Y)\leq(n/2)\sum_{i=1}^{n}(1/2)|X_i-Y_i|\leq(n/2)(\Delta-n)$ が成り立つことから、グラフ G の直径、すなわち $\max\{d(X,Y)\}$ は $n(\Delta-n)/2$ で押さえられる。また枝の長さの定義から、任意の枝 $\{X,Y\}\in\mathcal{E}$ に対して $d(X,Y)=l(\{X,Y\})$ が成り立つ。

次に coupling の推移 $(X,Y)\mapsto(X',Y')$ を $(X,Y)\mapsto(\phi(X,\Lambda),\phi(Y,\Lambda))$ と定義する。ただし、$\Lambda\in[1,n)$ は一様実数乱数で、ϕ は 2 節で定義した update function である。ここで任意の対 $\{X,Y\}\in\mathcal{E}$ に対して、

$$\mathrm{E}[d(X',Y')]\leq\beta d(X,Y),\ \beta=1-1/(n(n-1)^2),\tag{1}$$

を示そう。以下では $\{X,Y\}$ の支持対を $\{j_1,j_2\}$ で表す。また、一般性を失うことなく $j_1<j_2$ かつ $X_{j_2}+1=Y_{j_2}$ を仮定する。

Case 1: $\lfloor\Lambda\rfloor=j_2-1$ の場合について、

$$\mathrm{E}[d(X',Y')|\lfloor\Lambda\rfloor=j_2-1]\leq d(X,Y)-(1/2)(n-j_2+1)/(n-1)$$

を示す。$j_1=j_2-1$ の時、条件 1 より $X'=Y'$ である。ゆえに $d(X',Y')=0$ となる。以下、$j_1<j_2-1$ の場合について考える。いま $b'=X_{j_2-1}+X_{j_2}$ とし、$b''=Y_{j_2-1}+Y_{j_2}$ とする。この時 $X_{j_2}+1=Y_{j_2}$ より $b'+1=b''$ が成り立つ。マルコフ連鎖 $\mathcal{M}_{\mathrm{D}}$ の update function の定義から

$$\begin{aligned}X'_{j_2-1}=k &\Leftrightarrow [g_{b'}(k-1|u_{j_2-1},u_{j_2})\leq\Lambda-\lfloor\Lambda\rfloor<g_{b'}(k|u_{j_2-1},u_{j_2})]\\ Y'_{j_2-1}=k &\Leftrightarrow [g_{b'+1}(k-1|u_{j_2-1},u_{j_2})\leq\Lambda-\lfloor\Lambda\rfloor<g_{b'+1}(k|u_{j_2-1},u_{j_2})]\end{aligned}$$

を得る。前節の補題 3 で述べたように alternating inequalities

$$\begin{aligned}0 &= g_{b'+1}(0|u_{j_2-1},u_{j_2})=g_{b'}(0|u_{j_2-1},u_{j_2})\\ &\leq g_{b'+1}(1|u_{j_2-1},u_{j_2})\leq g_{b'}(1|u_{j_2-1},u_{j_2})\leq\cdots\\ &\leq g_{b'+1}(b'-1|u_{j_2-1},u_{j_2})\leq g_{b'}(b'-1|u_{j_2-1},u_{j_2})=g_{b'+1}(b'|u_{j_2-1},u_{j_2})=1\end{aligned}$$

が成り立つので、

$$\begin{pmatrix}X'_{j_2-1}\\ Y'_{j_2-1}\end{pmatrix}\in\left\{\begin{pmatrix}1\\1\end{pmatrix},\begin{pmatrix}1\\2\end{pmatrix},\begin{pmatrix}2\\2\end{pmatrix},\begin{pmatrix}2\\3\end{pmatrix},\ldots,\begin{pmatrix}b'-1\\b'-1\end{pmatrix},\begin{pmatrix}b'-1\\b'\end{pmatrix}\right\}$$

を得る。この時、もし $X'_{j_2-1}=Y'_{j_2-1}$ ならば、$\{X',Y'\}$ の支持対は $\{j_1,j_2\}$ となり、$d(X',Y')=d(X,Y)$ となる。また、もし $X'_{j_2-1}\neq Y'_{j_2-1}$ ならば、$\{X',Y'\}$ の支持対は $\{j_1,j_2-1\}$ となり、$d(X',Y')=d(X,Y)-(n-j_2+1)/(n-1)$ となる。

ここで $u_{j_2-1}\geq u_{j_2}$ の場合

$$\begin{aligned}&\Pr[X'_{j_2-1}\neq Y'_{j_2-1}|\lfloor\Lambda\rfloor=j_2-1]-\Pr[X'_{j_2-1}=Y'_{j_2-1}|\lfloor\Lambda\rfloor=j_2-1]\\ &= \textstyle\sum_{k=1}^{b'-1}[g_{b'}(k|u_{j_2-1},u_{j_2})-g_{b'+1}(k|u_{j_2-1},u_{j_2})]\\ &\quad -\textstyle\sum_{k=1}^{b'-1}[g_{b'+1}(k|u_{j_2-1},u_{j_2})-g_{b'}(k-1|u_{j_2-1},u_{j_2})]\geq 0\end{aligned}$$

が成り立つ。条件 1 から、$u_{j_2-1} \geq u_{j_2}$ は常に成り立ち

$$\begin{aligned}\Pr[X'_{j_2-1} = Y'_{j_2-1} | \lfloor \Lambda \rfloor = j_2 - 1] &\leq (1/2), \\ \Pr[X'_{j_2-1} \neq Y'_{j_2-1} | \lfloor \Lambda \rfloor = j_2 - 1] &\geq (1/2)\end{aligned}$$

を得る。ゆえに

$$\begin{aligned}\mathrm{E}[d(X', Y') | \lfloor \Lambda \rfloor = j_2 - 1] &\leq (1/2)d(X, Y) + (1/2)(d(X, Y) - (n - j_2 + 1)/(n - 1)) \\ &= d(X, Y) - (1/2)(n - j_2 + 1)/(n - 1)\end{aligned}$$

が成り立つ。

Case 2: $\lfloor \Lambda \rfloor = j_2$ の場合についても、Case 1 と同様に $\mathrm{E}[d(X', Y') | \lfloor \Lambda \rfloor = j_2] \leq d(X, Y) + (1/2)(n - j_2)/(n - 1)$ を示すことができる。

Case 3: $\lfloor \Lambda \rfloor \neq j_2 - 1$ かつ $\lfloor \Lambda \rfloor \neq j_2$ の場合。この時 $\{X', Y'\}$ の支持対 $\{j'_1, j'_2\}$ は $j_2 = \max\{j'_1, j'_2\}$ となる。したがって $d(X, Y) = d(X', Y')$ である。

それぞれの場合の生起する確率は Case 1 が $1/(n-1)$、Case 2 が高々$1/(n-1)$、Case 3 が残りである。以上のことから。

$$\begin{aligned}\mathrm{E}[d(X', Y')] &\leq d(X, Y) - \frac{1}{n-1}\frac{1}{2}\frac{n - j_2 + 1}{n-1} + \frac{1}{n-1}\frac{1}{2}\frac{n - j_2}{n-1} = d(X, Y) - \frac{1}{2(n-1)^2} \\ &\leq \left(1 - \frac{1}{2(n-1)^2}\frac{1}{\max_{\{X,Y\} \in \mathcal{E}}\{d(X, Y)\}}\right) d(X, Y) = \left(1 - \frac{1}{n(n-1)^2}\right) d(X, Y)\end{aligned}$$

を得る。

いま、$D \stackrel{\text{def.}}{=} d(X_{\mathrm{U}}, X_{\mathrm{L}})$ と $\tau_0 \stackrel{\text{def.}}{=} n(n-1)^2(1 + \ln D)$ を定義する。この時、

$$\begin{aligned}\Pr[T_* > \tau_0] &= \Pr\left[\Phi^0_{-\tau_0}(X_{\mathrm{U}}, \boldsymbol{\Lambda}) \neq \Phi^0_{-\tau_0}(X_{\mathrm{L}}, \boldsymbol{\Lambda})\right] = \Pr\left[\Phi^{\tau_0}_0(X_{\mathrm{U}}, \boldsymbol{\Lambda}) \neq \Phi^{\tau_0}_0(X_{\mathrm{L}}, \boldsymbol{\Lambda})\right] \\ &\leq \textstyle\sum_{(X,Y) \in \Omega^2} d(X, Y) \Pr\left[X = \Phi^{\tau_0}_0(X_{\mathrm{U}}, \boldsymbol{\Lambda}), Y = \Phi^{\tau_0}_0(X_{\mathrm{L}}, \boldsymbol{\Lambda})\right] \\ &= \mathrm{E}\left[d\left(\Phi^{\tau_0}_0(X_{\mathrm{U}}, \boldsymbol{\Lambda}), \Phi^{\tau_0}_0(X_{\mathrm{L}}, \boldsymbol{\Lambda})\right)\right] \leq \left(1 - \frac{1}{n(n-1)^2}\right)^{\tau_0} d(X_{\mathrm{U}}, X_{\mathrm{L}}) \\ &= \left(1 - \frac{1}{n(n-1)^2}\right)^{n(n-1)^2(1+\ln D)} D \leq \mathrm{e}^{-1}\mathrm{e}^{-\ln D} D \leq \frac{1}{\mathrm{e}}\end{aligned}$$

が得られる。ここで coalescence time の持つ submultiplicativity（[14]）を考慮すると、任意の $k \in \mathbb{Z}_+$ に対して、$\Pr(T_* > k\tau_0) \leq (\Pr(T_* > \tau_0))^k \leq (1/\mathrm{e})^k$ が得られる。したがって

$$\begin{aligned}\mathrm{E}[T_*] &= \textstyle\sum_{t=0}^{\infty} t \Pr[T_* = t] \leq \tau_0 + \tau_0 \Pr[T_* > \tau_0] + \tau_0 \Pr[T_* > 2\tau_0] + \cdots \\ &\leq \tau_0 + \tau_0/\mathrm{e} + \tau_0/\mathrm{e}^2 + \cdots = \tau_0/(1 - 1/\mathrm{e}) \leq 2\tau_0.\end{aligned}$$

となる。いま $n \leq \Delta$ より、明らかに $D \leq n(\Delta - n)/2 \leq \Delta^2$ である。ゆえに $\mathrm{E}[T_*] = \mathrm{O}(n^3 \ln \Delta)$ を得る。□

最後にアルゴリズム 1 の計算時間について議論する。

定理 2 の証明: マルコフ連鎖の coalescence time を T_* で表す。ここで T_* は確率変数である。いま $K = \lceil \log_2 T_* \rceil$ とする。アルゴリズム 1 は $(K+1)$ 回目の反復で T を -2^K とした時に終了する。したがってアルゴリズム 1 で生成する乱数の総数は $2^K \leq 2T_*$ で押さえられ、推移回数は $2(2^0 + 2^1 + 2^2 + \cdots + 2^K) < 2 \cdot 2 \cdot 2^K \leq 8T_*$ で押さえられる。各乱数は定数時間で得られると仮定するとマルコフ連鎖の推移は定数時間で行われる。また、アルゴリズム 1 の Step 4 “Coalescence check” に必要な時間は $\mathrm{O}(n)$ である。したがって合計計算時間は $\mathrm{O}(\mathrm{E}[2T_*] + \mathrm{E}[8T_*] + \mathrm{E}[K+1]n) = \mathrm{O}(\mathrm{E}[T_*]) = \mathrm{O}(n^3 \ln \Delta)$ となる。 □

なお条件 1 は $\mathrm{O}(n \ln n)$ で得られる。

参考文献

[1] Bubley, R., Dyer, M.: Path coupling: A technique for proving rapid mixing in Markov chains, *38th Annual Symposium on Foundations of Computer Science*, IEEE, San Alimitos, 1997, 223–231.

[2] Bubley, R.: *Randomized Algorithms : Approximation, Generation, and Counting*, Springer-Verlag, New York, 2001.

[3] Burr, T. L.: Quasi-equilibrium theory for the distribution of rare alleles in a subdivided population: justification and implications, *Ther. Popul. Biol.*, **57** (2000) 297–306.

[4] Burr, D., Doss, H., Cooke, G. E., Goldschmidt-Clermont, P. J.: A meta-analysis of studies on the association of the platelet PlA polymorphism of glycoprotein IIIa and risk of coronary heart disease, *Stat. Med.*, **22** (2003) 1741–1760.

[5] Dimakos, X. K.: A guide to exact simulation, *International Statistical Review*, 69 (2001), 27–48.

[6] Durbin, R., Eddy, R., Krogh, A., Mitchison, G.: *Biological sequence analysis: probabilistic models of proteins and nucleic acids*, Cambridge Univ. Press, 1998.

[7] Graham, J., Curran, J., Weir, B. S.: Conditional genotypic probabilities for microsatellite loci, *Genetics*, **155** (2000) 1973–1980.

[8] Kitada, S., Hayashi, T., Kishino, H.: Empirical Bayes procedure for estimating genetic distance between populations and effective population size, *Genetics*, **156** (2000) 2063–2079.

[9] Laval, G, SanCristobal, M., Chevalet C.: Maximum-likelihood and Markov chain monte carlo approaches to estimate inbreeding and effective size form allele frequency changes, *Genetics*, **164** (2003) 1189–1204.

[10] Matsui, T. and Kijima, S.,: Polynomial time perfect sampler for discretized Dirichlet distribution, METR 2003-17, Mathematical Engineering Technical Reports, University of Tokyo, 2003. (available from http://www.keisu.t.u-tokyo.ac.jp/Research/techrep.0.html)

[11] Matsui, T., Motoki, M., and Kamatani, N.,: Polynomial time approximate sampler for discretized Dirichlet distribution, *14th International Symposium on Algoeithms and Computation*, LNCS, **2906** (2003) 676–685.

[12] Niu, T., Qin, Z. S., Xu, X., Liu, J. S.: Bayesian haplotype inference for multiple linked single-nucleotide polymorphisms, *Am. J. Hum. Genet.*, **70** (2002) 157–169.

[13] Pritchard, J. K., Stephens, M., Donnely, P.: Inference of population structure using multilocus genotype data, *Genetics*, **155** (2000) 945–959.

[14] Propp, J. and Wilson, D.: Exact sampling with coupled Markov chains and applications to statistical mechanics, *Random Structures and Algorithms*, **9** (1996), 232–252.

[15] Propp, J. and Wilson, D.: How to get a perfectly random sample from a generic Markov chain and generate a random spanning tree of a directed graph, *J. Algorithms*, **27** (1998), 170–217.

[16] Robert, C. P.: *The Bayesian Choice*, Springer-Verlag, New York, 2001.

[17] Wilson, D.: How to couple from the past using a read-once source of randomness, *Random Structures and Algorithms*, **16** (2000), 85–113.