修士論文

パーフェクトサンプリング法の設計

東京大学 大学院情報理工学系研究科
数理情報学専攻

来嶋 秀治

指導教官　松井 知己 助教授

2004年2月2日

論文要旨

本論文では 2 つのサンプリング法を提案する。提案するサンプリング法はマルコフ連鎖を用いたサンプリング法で、Coupling From The Past (CFTP) に基づき定常分布に厳密に従うサンプリングを可能にする。提案するアルゴリズムはラスベガス型 (Las Vegas type) の乱択アルゴリズム (randomized algorithm) で、確率的に終了し厳密な解を返す。本論文で扱う問題は 2 行分割表の一様サンプリングと離散化 Dirichlet 分布に従うサンプリングである。それぞれの問題に対して収束の速い (rapidly mixing) マルコフ連鎖を提案する。さらに、これらのマルコフ連鎖が単調 (monotone) であることを示し、単調 CFTP (monotone CFTP) アルゴリズムを設計する。

目 次

第 1 章　序論 **1**

第 2 章　マルコフ連鎖の収束性に関連する既存の研究 **5**
2.1 Path Coupling 5
2.2 Coupling From The Past 7
2.3 Coalescence Time と Mixing Time 8

第 3 章　2 行分割表のパーフェクトサンプリング法 **11**
3.1 サンプリングアルゴリズム 11
3.2 マルコフ連鎖の単調性 12
3.3 アルゴリズムの計算時間 15

第 4 章　離散化 Dirichlet 分布に従うパーフェクトサンプリング法 **18**
4.1 サンプリングアルゴリズム 18
4.2 マルコフ連鎖の単調性 19
4.3 アルゴリズムの計算時間 23

第 5 章　結論 **29**

謝辞 **29**

参考文献 **30**

付 録 A Read Once アルゴリズム **35**

付 録 B 計算機実験 **37**
B.1 2 行分割表の Coalescence Time 37
B.2 離散化 Dirichlet 分布の Coalescence Time 41

第1章

序論

本論文では2つのサンプリング法を提案する。提案するサンプリング法はマルコフ連鎖を用いたサンプリング法で、Coupling From The Past (CFTP) に基づき定常分布に厳密に従うサンプリング (パーフェクトサンプリング: Perfect Sampling) を可能にする。提案するアルゴリズムはラスベガス型 (Las Vegas type) の乱択アルゴリズム (randomized algorithm) で、確率的に終了し厳密な解を返す。本論文で扱う問題は2行分割表の一様サンプリングと離散化 Dirichlet 分布に従うサンプリングである。それぞれの問題に対して収束の速い (rapidly mixing) マルコフ連鎖を提案する。さらに、これらのマルコフ連鎖が単調 (monotone) であることを示し、単調 CFTP (monotone CFTP) アルゴリズムを設計する。

マルコフ連鎖を用いたサンプリングは、マルコフ連鎖モンテカルロ (Markov chain Monte Carlo: MCMC) 法、Gibbs サンプラー、EM アルゴリズム、Metropolis-Hasting (M-H) シミュレーションといった形で、統計物理、医療統計、遺伝情報学、数値積分、画像処理、最適化算法、ネットワークシステムなど多岐にわたる分野に現れる。近年の計算機能力の向上により大規模な問題に対する計算が可能となり、これらの手法が実用的に用いられる場面が増えてきた。これに伴い、マルコフ連鎖を用いたランダム生成法に関する理論的研究が盛んである。中でもマルコフ連鎖の収束の速さ (mixing time) の算定について、組合せアルゴリズム理論の研究者を中心に精力的な研究が行われている [21, 24, 38, 48]。

一般にマルコフ連鎖を用いたサンプリングの用いられる対象は、標本空間が組合せ的に複雑で状態数が大きく、時には個数を知ることさえ困難である。このような対象に対して直接サンプリングは困難であり、棄却サンプリングは棄却される確率が非常に大きくなってしまい効率的でなくなる。そこで目標の分布を極限分布に持ち、収束の速い (rapidly mixing) マルコフ連鎖を設計することで効率的なサンプリングの実現を目指す。すなわち、目的の分布を極限分布に持つマルコフ連鎖を設計し、十分な回数の推移を行いランダムサンプルを生成する。したがってマルコフ連鎖の収束について推移の回数と極限分布からの分布の偏りの大きさの関係が議論される。

推移確率行列が陽に与えられたマルコフ連鎖に対しては、推移確率行列の固有値に注目することで mixing time の算定が可能である [45]。近年、Boyd et al は状態空間と推移グラフの与えられたマルコフ連鎖に対して SDP (半正定値計画) を解き、第2固有値の大きさを最小にする方法について議論している [5]。しかし状態空間の大きな対象、あるいはサイズを知ることさえ困難な対象に対して、推移確率行列を陽に与えられないため、固有値を直接算定する手法は今のところ現実的なものとはなっていない。

現在、マルコフ連鎖の mixing time の算定法としてよく用いられている手法はコンダクタンス (Conductance) 法と Coupling 法である。コンダクタンス法は Sinclair and Jerrum の提案した手法で、マルコフ連鎖の推移グラフ上で確率の流れを考え、コンダクタンス (流れにくさの量) を計算する手法である [46, 44]。一方、Coupling は古くからマルコフ連鎖の収束を示す道具として用いられていたが Aldous が 1983 年にこの手法を持ちいて mixing time を算定する手法を提案した [3]。また、Bubley and Dyer は Coupling 法を改良した Path Coupling 法を提案し、さらに使いやすい物にしている [7, 6]。一般的にコンダクタンス法と Coupling 法を比較すると、コンダクタンス法は多くのマルコフ連鎖に対

表 1.1. 分割表の一様サンプリングに関連する研究。

1985,	Diaconis and Effron [12],	2 元分割表の検定法の提案,
1995,	Diaconis and Saloff-Coste [13],	$m^* \times n^*$ 分割表の弱多項式時間近似一様生成法,
1997,	Dyer, Kannan and Mount, [18]	2 元分割表数え上げ問題の計算複雑度,
1998,	Hernek [23],	$2 \times n$ 分割表の擬多項式時間近似一様生成法,
2000,	Dyer and Greenhill [17],	$2 \times n$ 分割表の弱多項式時間近似一様生成法,
2002,	Cryan et al. [11],	$m^* \times n$ 分割表の弱多項式時間近似一様生成法,
2003,	Matsui, Matsui and Ono [33],	$2 \times \cdots \times 2 \times n$ 分割表の弱多項式時間近似一様生成法,
2003,	Kijima and Matsui [28],	$2 \times n$ 分割表の弱多項式時間厳密一様生成法,
		(注: m^*, n^* はそれぞれ定数であることを表す。)

して計算が行いやすい反面算定が緩いことが多いのに対して、Coupling 法は算定が困難だが算定できたならタイトな算定が行えると言われている [21]。

マルコフ連鎖を用いたサンプリングの多くは近似サンプリングである。しかし、近似サンプリングでは、どんなに推移を繰り返しても定常分布に厳密に従うサンプリングは不可能である。これに対し、Propp and Wilson は単純で驚くべきアルゴリズムを考案した [40]。そのアルゴリズムは CFTP アルゴリズム (あるいは backward coupling) と呼ばれ画期的アルゴリズムとして注目を浴びている [41, 14, 22, 50]。CFTP アルゴリズムは任意のエルゴード的な有限マルコフ連鎖に対して、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 (パーフェクトサンプリング) を可能とする。

パーフェクトサンプリングを行う利点を以下に 3 つ挙げる。1 つ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。多くの場合、サンプリングは近似アルゴリズムの一部として現れる。この時アルゴリズムの精度の議論を行うために、誤差パラメータの設定が必要ないことは算定を容易にする。また特に精度の高いサンプリングを要する時、パーフェクトサンプリングは近似サンプリングよりも速いアルゴリズムとなる。2 つ目に自動終了アルゴリズムであることが挙げられる。近似サンプリングでは、実際には算定が緩い場合でも理論的な精度の裏打ちされた結論を得るためには理論に基づいた回数の推移を行う必要がある。しかし CFTP に基づくパーフェクトサンプリングでは、アルゴリズムが終了すれば厳密解なので、算定よりもオーダーとして早く終わる可能性もある。3 つ目に、CFTP における coalescence time と mixing time の間の関係が知られており、実行可能なパーフェクトサンプリングが設計できれば、計算機実験を行い、実験的にマルコフ連鎖の収束スピードの実験的算定手法として価値がある。また、シミュレーションにおける収束性の指標としても、その可能性が模索されている。

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ある種の「単調性」(monotone) がある時、効率的な CFTP アルゴリズムの設計が可能となる [40, 41, 14, 22, 50]。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。

本論文では分割表の一様サンプリングと Dirichlet 分布に従うサンプリングのそれぞれに対して単調なマルコフ連鎖を提案し単調 CFTP アルゴリズムを設計する。

分割表は与えられた行和、列和を満たすような非負整数からなる表 (行列) で、統計データを扱う際に用いられる。分割表に対する統計学的興味として、行と列の独立性があげられる。Fischer の提案した正確検定はその一つの検定法である。Diaconis and Effron は分割表の独立性の検定の議論を行っている [12]。正確検定は分割表を全列挙することで行われるが、分割表の全列挙は困難である。実用的には、正確検定で現れる p 値の計算にマルコフ連鎖モンテカルロ (MCMC) 法が用いられる。

分割表の一様生成としてマルコフ連鎖を用いる手法に関する研究は数多く存在する。Diaconis and Saloff-Coste [13] は 2 元分割表に対する、素朴なマルコフ連鎖の収束性について議論している。彼らは分割表の行数、列数を固定した時、素朴なマルコフ連鎖が多項式時間で収束することを示した。Dyer, Kannan and Mount [18] は 2 元分割表の個数を数える問題が行数を 2 に固定した場合でさえ#P 完全であることを示した。#P 完全は数え上げ問題に対する計算量のクラスで、NP 困難のクラスに属し、その中でも比較的難しい問題のクラスとして認識されている [47]。彼らはこの数え上げ問題の近似解法を設計する中で別のマルコフ連鎖を提案している。周辺和が十分大きい時、彼らの提案したマルコフ連鎖は行数および列数の多項式時間で収束する。

2 行分割表に対して、Hernek [23] は素朴なマルコフ連鎖の mixing time が表の合計値と列数の多項式時間で押さえられることを示した。Hernek はマルコフ連鎖の mixing time を coupling theorem を用いて示している。Dyer and Greenhill [17] は 2 行分割表に対して、急速に収束 (rapidly mixing) する新しいマルコフ連鎖を提案した。そのマルコフ連鎖は表の合計値の対数と列数の多項式時間で収束する。彼らはそれを Bubley and Dyer [7] の提案した path coupling の手法を用いて示している。Matsui, Matsui and Ono[33] は Dyer and Greenhill の結果を $2 \times \cdots \times 2 \times J$ 分割表に拡張した。また、最近の研究では Cryan, Dyer, Goldberg, Jerrum and Martin [11] は行数 (または列数) 固定の時、Dyer and Greenhill のマルコフ連鎖を拡張した 2×2 連鎖が rapidly mixing であることをコンダクタンス法を用いて示している。現在、一般の $m \times n$ 分割表に対する rapidly mixing なマルコフ連鎖の存在については大きな未解決問題である。

本論文では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 Dyer and Greenhill のマルコフ連鎖の推移のいくつかを制限したものである。このマルコフ連鎖に対し、状態空間に特別な半順序を導入し、マルコフ連鎖の単調性を示した。また、提案するマルコフ連鎖の mixing time は $n(n-1)^2 \ln(nN/\varepsilon)$ で押さえられる。ただし n は列数、N は表の合計値をあらわす。この証明は Dyer and Greenhill と同様の path coupling 法を用いて行ったが、彼らの方法には無い前処理が必要であった。Propp and Wilson は [40] で、もしマルコフ連鎖が多項式時間の mixing time を持ち単調なら、coalescence tiem も多項式時間となることを示している。この技法に今回得られた結論を適用することで、2 行分割表に対する多項式時間のパーフェクトサンプリングアルゴリズムを設計した。提案したアルゴリズムの計算時間の期待値は $\mathrm{O}(n^3 \ln N)$ である。この結果は#P 完全な問題に対して、一様生成が数え上げに比べて易しいことを示唆する重要な結果といえる。

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 [42]。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 [39]。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。Dirichlet 分布は同系交配係数とアレル頻度交換の効果の大きさを見積もり [30]、また多形性と疾患の関係を調べるためのメタ解析を行うを行う際にも用いられる [9]。Burr は [8] で、細分された母集団中の希少アレルの分布に対する擬似均衡理論の検査に Dirichlet 分布を用いている。Kitada, Hayashi and Kishino は母集団と有効な母集団の大きさとの間の遺伝的距離の推定に Dieichlet 分布を用いている [29]。Graham, Curran, and Weir は法医学のためのマイクロサテライト座位に対する条件付遺伝子型確率の見積もりに Dirichlet 分布を用いている [20]。

(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 ([15] 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 [32]。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。

本論文で提案するマルコフ連鎖は Matsui, Motoki and Kamatani のマルコフ連鎖の推移のいくつかを制限したもので、各推移では離散化ベータ分布 (2 次元の Dirichlet 分布) に従う確率変数の生成を行う。また、マルコフ連鎖の状態空間に 2 行分割表の場合と同様の半順序関係を導入する。しかし、推移確率が一様でないため、単調性の証明は分割表の場合のように簡単ではなく 'alternating inequality' を導入して示した。提案するマルコフ連鎖の各推移の計算量は [32] 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。

本論文の構成は以下のようになっている。次の章でマルコフ連鎖の収束性に関連する研究のうち、本論文に直接関係する 2 つの手法について述べる。第 3 章では $2\times n$ 分割表に対する新しいマルコフ連鎖を提案し、CFTP アルゴリズムに基づいたサンプリング法を提案する。第 4 章では離散化 Dirichlet 分布に従う新しいマルコフ連鎖を提案し、CFTP アルゴリズムに基づいたサンプリング法を提案する。第 5 章で本論文をまとめる。

第2章

マルコフ連鎖の収束性に関連する既存の研究

本章では、マルコフ連鎖の収束性に関連する既往の手法について述べる。第 1 節では Bubley and Dyer の提案した Path Coupling について述べ、第 2 節で Propp and Wilson の提案した CFTP について述べる。第 3 節では近似サンプリングにおける mixing time と CFTP における coalescence time の関係について述べる。

2.1 Path Coupling

本論文では実数全体の集合を $\mathbb{R}$ で表し、また整数 (非負整数、正整数) 全体の集合をそれぞれ $\mathbb{Z}$ ($\mathbb{Z}_+$, $\mathbb{Z}_{++}$) で表す。有限の状態空間 Ω と推移確率行列 P をもつマルコフ連鎖 $\mathcal{M}$ を考える。マルコフ連鎖 $\mathcal{M}$ が既約 (*irreducible*) であるとは任意の 2 つの状態 $\forall x, \forall y \in \Omega$ に対して、$\exists t > 0$, $\Pr(X^t = y \mid X^0 = x) > 0$ であることを言う。また、マルコフ連鎖 $\mathcal{M}$ が非周期的 (*aperiodic*) であるとは $\forall x \in \Omega$, $\gcd\{t \in \mathbb{Z}_{++} \mid \Pr(X^t = x \mid X^0 = x) > 0\} = 1$ を言う。ただし gcd は最大公約数を表す。既約で非周期的な有限マルコフ連鎖をエルゴード的 (*ergodic*) と呼び、唯一の定常分布を持ち、極限分布は定常分布に一致する。

いま、エルゴード的なマルコフ連鎖 $\mathcal{M}$ は定常分布 π を持つとする。マルコフ連鎖 $\mathcal{M}$ が *reversible* であるとは次の等式が成り立つことを言う。

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

ただし $P(x,y)$ は x から y への推移確率を表す。このことは定数 $c > 0$ に対して、関数 $q(x) = c\pi(x)$ が

$$\forall x, y \in \Omega,\ q(x)P(x,y) = q(y)P(y,x).$$

を満たすことと等価である。上の等式を *detailed balance equation* という。マルコフ連鎖が detailed balance equation を満たす時、M-H 連鎖とも言う。

同一の有限状態空間 Ω 上の 2 つの確率分布 ν_1 と ν_2 が与えられた時、ν_1 と ν_2 の間の総分布距離 (*total variation distance*) は

$$\mathrm{d_{TV}}(\nu_1, \nu_2) \stackrel{\text{def.}}{=} \max_{A \subseteq \Omega} \left\{ \sum_{x \in A} (\nu_1(x) - \nu_2(x)) \right\} \equiv \frac{1}{2} \sum_{x \in \Omega} |\nu_1(x) - \nu_2(x)|$$

と定義される。状態空間 Ω を持つエルゴード的なマルコフ連鎖 $\mathcal{M}$ に対して *mixing time* は任意の正数 $\varepsilon < 1$ に対して

$$\tau(\varepsilon) \stackrel{\text{def.}}{=} \max_{x \in \Omega} \{\min\{t \mid \forall s \geq t,\ \mathrm{d_{TV}}(\pi, P_x^s) \leq \varepsilon\}\}$$

と定義される。ただし、π はマルコフ連鎖 $\mathcal{M}$ の定常分布とし、P_x^s は初期状態 $x \in \Omega$ として時刻 0 からの時刻 $s \geq 0$ まで推移させた時のマルコフ連鎖 $\mathcal{M}$ の確率分布とする。また特に $\tau = \tau(1/\mathrm{e})$ を *mixing rate* と呼ぶ。

次の Coupling Lemma は古くから知られた補題で、エルゴード的なマルコフ連鎖が唯一の定常分布に収束することを示す際に用いられる [3, 6, 22, 21]。

定理 2.1 (Coupling Lemma) 確率変数 X, Y は同一の状態空間 Ω にあり、それぞれ確率分布 μ, ν にしたがう。この時 $\mathrm{d_{TV}}(\mu,\nu) \leq \Pr(X \neq Y) = \sum_{x\neq y} \Pr(X=x, Y=y)$ が成り立つ。

証明: 部分集合 $A^* \subseteq \Omega$ は $\max_{A\subseteq\Omega} \left\{\sum_{s\in A}(\mu(s)-\nu(s))\right\}$ を実現する A とする。この時、

$$\begin{aligned}\mathrm{d_{TV}}(\mu,\nu) &= \max_{A\subseteq\Omega}\left\{\sum_{s\in A}(\mu(s)-\nu(s))\right\} = \sum_{s\in A^*}(\Pr(X=s)-\Pr(Y=s)) \\ &\leq \sum_{s\in A^*}\Pr(X=s, Y\neq s) \leq \sum_{s\in\Omega}\Pr(X=s, Y\neq s) = \Pr(X\neq Y)\end{aligned}$$

より題意を得る。 □

次の Path Coupling 定理は mixing time の上限を算定するための効果的な手法のひとつである。下の定理は Bubley and Dyer ([7]) のオリジナルのものに対して、枝の長さの整数性を仮定していない点が異なるが、最小距離 d を導入することで同様の証明が行えこの変形は本質的なものではない。

定理 2.2 (Path Coupling [7]) 状態空間 Ω を持つマルコフ連鎖 $\mathcal{M}$ はエルゴード的とする。単純無向グラフ $G=(\Omega,\mathcal{E})$ は頂点集合 Ω と枝集合 $\mathcal{E} \subseteq \binom{\Omega}{2}$ を持ち、連結なグラフとする。いま、枝の長さを $l : \mathcal{E} \to \mathbb{R}_{++}$ とする。グラフ G の任意の頂点間 $\{x,y\}$ に対して x と y の間の距離を $d(x,y)$ あるいは $d(y,x)$ で表し、x と y のグラフ G 上の最短パスの長さとする。ただしパスの長さとはパス上の枝の距離の合計である。いま coupling (X,Y) に対して coupling の推移 $(X,Y) \mapsto (X',Y')$ が存在して、X および Y に関する周辺推移確率が $\mathcal{M}$ に従い

$$0 < \exists\beta < 1,\ \forall\{X,Y\} \in \mathcal{E},\ \mathrm{E}[d(X',Y')] \leq \beta d(X,Y).$$

を満たすとする。この時、マルコフ連鎖 $\mathcal{M}$ の mixing time $\tau(\varepsilon)$ は $\tau(\varepsilon) \leq (1-\beta)^{-1}\ln(D/d\varepsilon)$ を満たす。ただし、$d \stackrel{\text{def.}}{=} \min\{d(x,y) \mid \forall x,\forall y \in \Omega\}$ とし $D \stackrel{\text{def.}}{=} \max\{d(x,y) \mid \forall x,\forall y \in \Omega\}$ とする。

証明: 時刻 $t \geq 0$ において $X = X^t$ は分布 μ^t に従い、$Y = Y^t$ は定常分布 π に従っているとする。Coupling Lemma より、任意の時刻 t に対して、

$$\begin{aligned}\mathrm{d_{TV}}(\mu^t,\pi) &\leq \Pr(X^t \neq Y^t) = \sum_{x\neq y}\Pr(X^t=x, Y^t=y) \\ &\leq \sum_{x,y}\frac{d(x,y)}{d}\Pr(X^t=x,Y^t=y) = \frac{\mathrm{E}[d(X^t,Y^t)]}{d} \qquad (2.1)\end{aligned}$$

である。いま、任意の $X,Y \in \Omega$ に対して、要素列 $X = Z_1, Z_2, \ldots, Z_p = Y$ $(Z_i \in \Omega)$ が存在して $d(X,Y) = \sum_{i=1}^{p-1} d(Z_i, Z_{i+1})$ とする。定理の仮定から

$$\mathrm{E}[d(X',Y')] \leq \mathrm{E}[\textstyle\sum_{i=1}^{p-1} d(Z_i',Z_{i+1}')] \leq \sum_{i=1}^{p-1}\mathrm{E}[d(Z_i',Z_{i+1}')] \leq \beta d(Z_i,Z_{i+1}) = \beta d(X,Y)$$

が得られる。ここで (2.1) 式を変形し、$t = -\ln\beta\,\ln(D/d\varepsilon) \leq \tau(\varepsilon) = (1-\beta)^{-1}\ln(D/d\varepsilon)$ を代入すると

$$\frac{\mathrm{E}[d(X^t,Y^t)]}{d} \leq \beta\frac{d(X^{t-1},Y^{t-1})}{d} \leq \beta^t\frac{d(X^0,Y^0)}{d} \leq \beta^t\frac{D}{d} \leq -\mathrm{e}^{\ln\frac{D}{d\varepsilon}}\frac{D}{d} = \varepsilon$$

が成り立つ。 □

2.2 Coupling From The Past

本節では 1996 年に Propp and Wilson の提案した Coupling From The Past (CFTP) について述べる。CFTP のアイデアは以下のとおりである。エルゴード的なマルコフ連鎖を無限回推移させることで、定常分布に厳密に従うランダムサンプルを得ることができる。いま、仮想的に無限の過去から無限回推移しているマルコフ連鎖が存在するならば、そのマルコフ連鎖の現在の状態は定常分布に厳密に従うランダムサンプルである。マルコフの連鎖の現在の状態が、過去のある時点の状態に依らず、現時点における唯一実現可能な状態であるという証拠が得られたなら、その状態は定常分布に従う確率標本である。これが CFTP の発想である。

有限の状態空間 Ω をもち、推移確率行列 P をもつマルコフ連鎖 $\mathcal{M}$ は、既約で非周期、すなわちエルゴード的とする。$\mathcal{M}$ の推移規則は、一様実数乱数 $\lambda \in [0,1)$ が与えられた時、*update function* と呼ばれる決定的関数 $\phi : \Omega \times [0,1) \to \Omega$ によって記述される。ここで update function は $\forall x, y \in \Omega,\ P(x,y) = \Pr(\phi(x,\Lambda) = y)$ が満たされていなければ成らない。乱数列 $\boldsymbol{\lambda} = (\lambda[t_1], \lambda[t_1+1], \ldots, \lambda[t_2-1]) \in [0,1)^{t_2-t_1}$ が与えられた時の時刻 t_1 から t_2 への $\mathcal{M}$ の推移も決定的関数 $\Phi_{t_1}^{t_2}(x, \boldsymbol{\lambda}) : \Omega \times [0,1)^{t_2-t_1} \to \Omega$ で表現される。ただし、時刻 $t_1, t_2\ (t_1 < t_2)$ に対して $\Phi_{t_1}^{t_2}(x, \boldsymbol{\lambda}) \stackrel{\text{def.}}{=} \phi(\phi(\cdots(\phi(x, \lambda[t_1]), \ldots, \lambda[t_2-2]), \lambda[t_2-1])$ と定義する。

これらの準備を用いて、標準的な CFTP アルゴリズムは以下のように記述される。

アルゴリズム 1 (CFTP)

Step 1. シミュレーションの開始時刻を $T := -1$ とする。空列 $\boldsymbol{\lambda}$ を用意する。

Step 2. 乱数 $\lambda[T], \lambda[T+1], \ldots, \lambda[\lceil T/2 \rceil - 1]$ を生成し、数列 $\boldsymbol{\lambda}$ の先頭に挿入する。すなわち、$\boldsymbol{\lambda} := (\lambda[T], \lambda[T+1], \ldots, \lambda[-1])$ とする。

Step 3. Ω の全ての状態について、乱数列 $\boldsymbol{\lambda}$ を用いて時刻 T から時刻 0 までマルコフ連鎖を推移させる。

(a) もし $\exists y \in \Omega, \forall x \in \Omega,\ y = \Phi_T^0(x, \boldsymbol{\lambda})$ ならば y を返し、停止する。

(b) そうでなければ、シミュレーションの開始時刻を $T := 2T$ としてステップ 2 に戻る。

定理 2.3 (CFTP [40, 14]) 有限マルコフ連鎖 $\mathcal{M}$ はエルゴード的で状態空間 Ω をもち、update function $\phi : \Omega \times [0,1) \to \Omega$ で定義されているとする。この時、CFTP アルゴリズム (アルゴリズム 1) が確率 1 で終了するなら、得られる値は $\mathcal{M}$ の定常分布に厳密に従う確率変数の実現値である。

証明: アルゴリズム 1 が終了したとする。得られた値を y とし、終了した時の反復のシミュレーションの初期時刻を $T < 0$、推移に使う乱数の列を $\boldsymbol{\lambda} = (\lambda[T], \lambda[T+1], \ldots, \lambda[-1])$ とする。いま、一様乱数 $\lambda[s']\ (t \le \forall s' < T)$ を生成し、$\boldsymbol{\lambda}' = (\lambda[t], \lambda[t+1], \ldots, \lambda[T-1], \lambda[T], \lambda[T+1], \ldots, \lambda[-1])$ とする。ここで、$\lambda[s]\ (T \le s < 0)$ はアルゴリズムで得られた乱数であることに注意する。この時、このマルコフ連鎖は時刻 t の状態 $x \in \Omega$ に対して時刻 0 の状態 $\Phi_t^0(x, \boldsymbol{\lambda}')$ は P_x^t に従うランダムサンプルである。マルコフ連鎖 $\mathcal{M}$ はエルゴード的なので、$\forall \varepsilon, \exists t < T, \forall x \in \Omega, \mathrm{d_{TV}}(P_x^t, \pi) \le \varepsilon$ が成り立つ。一方、CFTP アルゴリズムで得られるサンプルは $\lambda[s']\ (t \le \forall s' < T)$ によらず、$y = \Phi_T^0(x, \boldsymbol{\lambda}) = \Phi_t^0(x, \boldsymbol{\lambda}')$ となる。よって題意を得る。 □

命題 2.4 有限の状態空間を持ち、推移確率の与えられた任意のエルゴード的なマルコフ連鎖に対して、確率 1 で収束するような update function が存在する。 ■

アルゴリズム 1 に対して、*coalescence time* T_* を $T_* \stackrel{\text{def.}}{=} \min\{t > 0 \mid -t \ge T,\ \exists y \in \Omega,\ \forall x \in \Omega,\ y = \Phi_{-t}^0(x, \boldsymbol{\lambda}^t)\}$ と定義する。ただし $\boldsymbol{\lambda}^t = (\lambda[-t], \lambda[-t+1], \ldots, \lambda[-1])$ はアルゴリズム 1 から得られる数列である。明らかに T_* は確率変数である。また、アルゴリズム 1 の中では coalescence time は陽に求まらないことに注意が必要である。

定理 2.3 より、無限時間のシミュレーション結果が (確率的に) 有限時間のアルゴリズムで得られる。しかし、このアルゴリズムでは Ω の状態数回のシミュレーションを行う必要があり、状態数の多い Ω に対しては実現が困難である。ところがもし、マルコフ連鎖にある意味で単調性が存在するならば、この問題は解決する。いま状態空間 Ω の要素全体に対して、半順序 $\succeq$ が存在すると仮定する。もし、全ての $x \succeq y$ $(x, y \in \Omega)$ に対して、$\phi(x, \lambda) \succeq \phi(y, \lambda)$ がどんな乱数 λ が与えられた時でも成り立つならば、その推移規則は単調であると言うことにする。また、そのようなマルコフ連鎖に対しても単調であると言うことにする。この時次の定理が成り立つ。

定理 2.5 (単調 CFTP [40, 14]) update function ϕ で定義されたマルコフ連鎖は半順序集合 $(\Omega, \succeq)$ に関して単調であり、$\exists x_{\max}, \exists x_{\min} \in \Omega$, $\forall x \in \Omega$, $x_{\max} \succeq x \succeq x_{\min}$ とする。この時 CFTP アルゴリズム (アルゴリズム 1) は確率 1 で終了し、$\exists y \in \Omega$, $\forall x \in \Omega$, $y = \Phi_T^0(x, \boldsymbol{\lambda})$ と $\Phi_T^0(x_{\max}, \boldsymbol{\lambda}) = \Phi_T^0(x_{\min}, \boldsymbol{\lambda})$ は等価である。 ■

マルコフ連鎖が定理 2.5 の条件を満たす時、アルゴリズム 1 の Step 4 (a) を

Step 4. (a)′ もし $\exists y \in \Omega$, $y = \Phi_T^0(x_{\max}, \boldsymbol{\lambda}) = \Phi_T^0(x_{\min}, \boldsymbol{\lambda})$ ならば y を返し、停止する。

で置き換えることができる。このアルゴリズムを monotne CFTP アルゴリズムと呼ぶ。

2.3 Coalescence Time と Mixing Time

CFTP アルゴリズムにおける coalescence time と近似サンプリングにおいてマルコフ連鎖の収束の速さの指標となる mixing time との間には関係があることが知られている [40, 14]。

まず、以下の議論を容易にするために、CTTF (Coupling To The Future) を定義する。これは CFTP がシミュレーションの初期時刻を過去に遡って行ったのに対して、終了時刻を未来に進めていくシミュレーションである。すなわち CTTF とは Ω の全ての状態について、共通の乱数列 $\boldsymbol{\lambda} = (\lambda[0], \lambda[1], \ldots, \lambda[T-1])$ を用いて時刻 0 から時刻 T までマルコフ連鎖を推移させることを言う。CTTF の colaesce を $\exists y \in \Omega$, $\forall x \in \Omega$, $y = \Phi_0^T(x, \boldsymbol{\lambda})$ と定義し、coalescence timeT^* は $T^* \stackrel{\text{def.}}{=} \min\{t > 0 \mid \Phi_0^t(x, \boldsymbol{\lambda})\}$ と定義する。

この時、次の命題が成り立つ [40]。

命題 2.6 CFTP アルゴリズムの coalescence time T_* と CTTF の coalescence time T^* は確率変数である。確率変数 T^* と T_* に対して次のことが成り立つ。

$$\forall t > 0,\ \Pr(T_* > t) = \Pr(T^* > t).$$

証明: 数列 $\boldsymbol{\lambda} \in [0,1)^t$ に対して $\exists y \in \Omega,\ \forall x \in \Omega,\ y = \Phi_{-t}^0(x, \boldsymbol{\lambda})$ が成り立つ時、同一の $\boldsymbol{\lambda}$ に対して $\exists y \in \Omega,\ \forall x \in \Omega,\ y = \Phi_0^t(x, \boldsymbol{\lambda})$ も成り立つ。確率変数列 $\boldsymbol{\Lambda} \in [0,1)^t$ は一様実数乱数の列とすると、

$$\begin{aligned}\Pr(T_* > t) &= 1 - \Pr(T_* \le t) = 1 - \Pr(\boldsymbol{\Lambda} \in \{\boldsymbol{\lambda} \in [0,1)^t \mid \exists y \in \Omega,\ \forall x \in \Omega,\ y = \Phi_{-t}^0(x, \boldsymbol{\lambda})\}) \\ &= 1 - \Pr(\boldsymbol{\Lambda} \in \{\boldsymbol{\lambda} \in [0,1)^t \mid \exists y \in \Omega,\ \forall x \in \Omega,\ y = \Phi_0^t(x, \boldsymbol{\lambda})\}) = 1 - \Pr(T^* \le t) = \Pr(T^* > t)\end{aligned}$$

となり、題意が示される。 □

系 2.7 確率変数 T^* と T_* に対して $\mathrm{E}[T_*] = \mathrm{E}[T^*]$ が成り立つ。 ■

次に、単調なマルコフ連鎖について、mixing time から coalescence time の期待値が算定できることを示す [40]。

定理 2.8 いま、状態空間 Ω をもち、update function ϕ で定義されるマルコフ連鎖 $\mathcal{M}$ は単調とし、定常分布を π、mixing rate を τ とする。マルコフ連鎖の半順序つき状態空間 $(\Omega, \succeq)$ 中の最大元と最小元をそれぞれ $x_{\max}$ および $x_{\min}$ で表す。ここで半順序つき状態空間中に「高さ」h を定義する。状態 x の

高さを $h(x)$ であらわし、任意の $x \succ y$ に対して $h(x) > h(y)$ とする。また、$h(x_{\min}) = 0$、$h(x_{\max}) = D$ とし、$d \stackrel{\text{def.}}{=} \min\{|h(x) - h(y)| \mid \forall (x,y) \in \Omega^2,\ h(x) \neq h(y)\}$ とする。この時、CFTP の coalescence time T_* は $\mathrm{E}(T_*) \leq 2\tau(1 + \ln(2D/d))$ を満たす。

証明: いま、初期状態 $x_{\min}$ から k 回マルコフ連鎖を推移させた状態を X_0^k であらわし、初期状態 $x_{\max}$ から k 回マルコフ連鎖を推移させた状態を X_1^k であらわす。ここで、マルコフ連鎖の単調性より $X_1^k \succeq X_0^k$ が成り立つ。この時、

$$
\begin{aligned}
\Pr(T^* > k) &= \Pr(X_0^k \neq X_1^k) \\
&\leq \sum_{(x,y)\in\Omega^2} \Pr(X_0^k = x, X_1^k = y) \frac{h(y) - h(x)}{d} \\
&= \frac{1}{d} \sum_{x\in\Omega} \sum_{y\in\Omega} \Big[\Pr(X_0^k = x, X_1^k = y) h(y) - \Pr(X_0^k = x, X_1^k = y) h(x) \Big] \\
&= \frac{1}{d} \left(\sum_{x\in\Omega} \sum_{y\in\Omega} \Pr(X_0^k = x, X_1^k = y) h(y) - \sum_{x\in\Omega} \sum_{y\in\Omega} \Pr(X_0^k = x, X_1^k = y) h(x) \right) \\
&= \frac{1}{d} \left(\sum_{y\in\Omega} \Pr(X_1^k = y) h(y) - \sum_{x\in\Omega} \Pr(X_0^k = x) h(x) \right) \\
&= \frac{1}{d} \left(\sum_{x\in\Omega} P_{x_{\max}}^k(x) h(x) - \sum_{x\in\Omega} P_{x_{\min}}^k(x) h(x) \right) \\
&= \frac{1}{d} \left(\sum_{x\in\Omega} (P_{x_{\max}}^k(x) - \pi(x)) h(x) - \sum_{x\in\Omega} (P_{x_{\min}}^k(x) - \pi(x)) h(x) \right) \\
&\leq \frac{1}{d} \left(\max_{A\subseteq\Omega} \left\{ \sum_{x\in A} (P_{x_{\max}}^k(x) - \pi(x)) h(x) \right\} + \max_{A\subseteq\Omega} \left\{ \sum_{x\in A} (\pi(x) - P_{x_{\min}}^k(x)) h(x) \right\} \right) \\
&\leq \frac{D}{d} \left(\max_{A\subseteq\Omega} \left\{ \sum_{x\in A} (P_{x_{\max}}^k(x) - \pi(x)) \right\} + \max_{A\subseteq\Omega} \left\{ \sum_{x\in A} (\pi(x) - P_{x_{\min}}^k(x)) \right\} \right) \\
&\leq \frac{2D}{d} \max_{x\in\Omega} \Big\{ \mathrm{d_{TV}}(\pi, P_x^k) \Big\}
\end{aligned}
$$

が成り立つ。ここで $\varepsilon = \Pr(T^* > k)$ とすると、劣乗法性[1] (*submultiplicativity* [40]) より、$\Pr(T^* > \alpha k) \leq \varepsilon^\alpha$ を得る。ゆえに

$$
\begin{aligned}
\mathrm{E}(T^*) &= \textstyle\sum_0^{+\infty} t \Pr(T^* = t) \\
&\leq k + \textstyle\sum_k^{+\infty} (t-k) \Pr(T^* = t) \\
&\leq k + k\Pr(T^* > t) + \textstyle\sum_{2k}^{+\infty} (t-2k) \Pr(T^* = t) \\
&\leq k + k\Pr(T^* > t) + k\Pr(T^* > 2t) + \textstyle\sum_{3k}^{+\infty} (t-3k) \Pr(T^* = t) \\
&\leq k + k\varepsilon + k\varepsilon^2 + \cdots \\
&= \frac{k}{\Pr(T^* \leq k)}
\end{aligned}
$$

を得る。マルコフ連鎖 $\mathcal{M}$ の mixing rate を τ とし、$k = \tau(1 + \ln \frac{2D}{d})$ とすると

$$
\Pr(T^* > k) \leq \frac{2D}{d} \max_x \Big\{ \mathrm{d_{TV}}(\pi, P_x^k) \Big\} \leq \frac{2D}{d} \frac{1}{\frac{2D}{d}\mathrm{e}} \leq \frac{1}{\mathrm{e}}
$$

[1] $\Pr(T^* > K_1 + K_2) \leq \Pr(T^* > K_1) \Pr(T^* > K_2)$

を得る。したがって

$$\mathrm{E}(T_*) = \mathrm{E}(T^*) \leq \frac{k}{\Pr(T^* \leq k)} \leq \frac{k}{1 - \frac{1}{\mathrm{e}}} < 2k = 2\tau \left(1 + \ln \frac{2D}{d}\right)$$

を得る。 □

最後にマルコフ連鎖の coalescence time の期待値から mixing time を算定する定理を述べる。この定理について述べている文献はこれまでに無いが、Propp and Wilson の結果から導くことができる。

定理 2.9 有限の状態空間 Ω をもち、推移確率 P をもつ単調なマルコフ連鎖 $\mathcal{M}$ の定常分布を π であらわす。マルコフ連鎖 $\mathcal{M}$ の coalescence time の期待値 $\mathrm{E}[T_*]$ が与えられた時、マルコフ連鎖の mixing time は $0 < \forall\varepsilon < 1$, $\tau(\varepsilon) \leq \mathrm{E}[T_*]\varepsilon^{-1}$ で抑えられる。

証明: いま 2 つの確率変数 X^k と Y^k は、それぞれ初期状態 x から k 回マルコフ連鎖を推移させた状態とする。定理 2.1 (Coupling Lemma) から、マルコフ連鎖が単調であることに注意して、$\forall k \in \mathbb{Z}_+$,

$$\mathrm{d_{TV}}(P_x^k, \pi) \leq \Pr[X^k \neq Y^k] \leq \Pr[X_0^k \neq X_1^k] = \Pr[T^* > k] = \Pr[T_* > k]$$

である。ただし、X_0^k および X_1^k は、それぞれ初期状態 $x_{\min}$ および $x_{\max}$ から CTTF を行ったときの時刻 k における状態である。T_* の非負性から、

$$k\Pr[T_* > k] = k \sum_{t=k+1}^{\infty} \Pr[T_* = t] \leq \sum_{t=1}^{\infty} t\Pr[T_* = t] = \mathrm{E}[T_*]$$

が成り立つので、

$$\mathrm{d_{TV}}(P_x^k, \pi) \leq \Pr[T_* > k] \leq \frac{\mathrm{E}[T_*]}{k}$$

である。今、$\tau(\varepsilon) \stackrel{\text{def.}}{=} \mathrm{E}[T_*]\varepsilon^{-1}$ とすると、

$$\mathrm{d_{TV}}(P_x^{\tau(\varepsilon)}, \pi) \leq \frac{E[T_*]}{E[T_*]\varepsilon^{-1}} = \varepsilon$$

となり題意を得る。 □

第3章

2行分割表のパーフェクトサンプリング法

本章では与えられた周辺和を満たす 2 行分割表を一様分布に厳密に従うサンプリングについて述べる。

3.1 サンプリングアルゴリズム

2 つのベクトル $\boldsymbol{r} = (r_1, r_2) \in \mathbb{Z}_{++}^2$ と、$\boldsymbol{s} = (s_1, \ldots, s_n) \in \mathbb{Z}_{++}^n$ は $\sum_{i=1}^2 r_i = \sum_{j=1}^n s_j = N \in \mathbb{Z}_{++}$ を満たすとする。行和および列和ベクトル $(\boldsymbol{r}, \boldsymbol{s})$ を持つ $2 \times n$ 分割表の集合 Ξ を

$$\Xi \stackrel{\text{def.}}{=} \left\{ X \in \mathbb{Z}_+^{2\times n} \;\middle|\; \begin{array}{l} \sum_{j=1}^n X[i,j] = r_i \ (1 \le \forall i \le 2), \\ \sum_{i=1}^2 X[i,j] = s_j \ (1 \le \forall j \le n) \end{array} \right\}$$

と定義する。ただし、$X[i,j]$ は i 行 j 列の値を表す。

以下で、$\boldsymbol{r}$ と $\boldsymbol{s}$ に対する Ξ を状態空間として持つ新しいマルコフ連鎖 $\mathcal{M}_{\mathrm{C}}$ を提案する。任意の行の添え字 $j \in \{1, \ldots, n-1\}$ に対して、

$$a_X(j) \stackrel{\text{def.}}{=} X[1,j] + X[1,j+1], \tag{3.1}$$

$$b_X(j) \stackrel{\text{def.}}{=} X[2,j] + X[2,j+1], \tag{3.2}$$

$$\theta_X(j) \stackrel{\text{def.}}{=} \min\{a_X(j), b_X(j), s_j, s_{j+1}\} + 1. \tag{3.3}$$

を定義する。マルコフ連鎖 $\mathcal{M}_{\mathrm{C}}$ の推移は次の update function $\phi : \Xi \times [1,n) \to \Xi$ によって定義される。ある時刻の状態 $X \in \Xi$ に対して、次の時刻の状態 $X' = \phi(X, \lambda) \in \Xi$ は実数乱数 $\lambda \in [1,n)$ によって次のように決まる。

$$X'[1,j] = \begin{cases} \min\{a_X(j), s_j\} - \lfloor (\lambda - \lfloor \lambda \rfloor)\ \theta_X(j) \rfloor & (j = \lfloor \lambda \rfloor), \\ a_X(j) - X'[1, \lfloor \lambda \rfloor] = a_X(j) - \min\{a_X(j), s_j\} + \lfloor (\lambda - \lfloor \lambda \rfloor)\ \theta_X(j) \rfloor & (j = \lfloor \lambda \rfloor + 1), \\ X[1,j] & (\text{otherwise}), \end{cases}$$

$$X'[2,j] = s_j - X'[1,j].$$

このマルコフ連鎖 $\mathcal{M}_{\mathrm{C}}$ は Dyer and Greenhill の提案したマルコフ連鎖 ([17]) を改造したものである。明らかに $\mathcal{M}_{\mathrm{C}}$ は有限で既約で非周期的なのでエルゴード的である。このマルコフ連鎖は一様分布を唯一の定常分布としてもつ。

さて、2 つの特別な分割表 $X_{\mathrm{U}}, X_{\mathrm{L}} \in \Xi$ を

$$X_{\mathrm{U}} \stackrel{\text{def.}}{=} \left(X[i,j] \in \mathbb{Z}_+ \;\middle|\; \exists k \in \{1, \ldots, n\},\ r_1 = \textstyle\sum_{j=1}^k X[1,j] \le \sum_{j=1}^k s_j,\ X[2,j] = 0\ (j < k) \right),$$

$$X_{\mathrm{L}} \stackrel{\text{def.}}{=} \left(X[i,j] \in \mathbb{Z}_+ \;\middle|\; \exists l \in \{1, \ldots, n\},\ r_1 = \textstyle\sum_{j=l}^n X[1,j] \le \sum_{j=l}^n s_j,\ X[2,j] = 0\ (j > l) \right).$$

と定義する。ここで $X_{\mathrm{U}}, X_{\mathrm{L}}$ はそれぞれ北西隅の規則および北東隅の規則で得られる表であることを付記する。

次のアルゴリズムが本論文で提案するサンプリング法である。

5	4	3	0	0	0	12
0	0	0	7	5	6	18
5	4	3	7	5	6	30

X_{U}

0	0	0	1	5	6	12
5	4	3	6	0	0	18
5	4	3	7	5	6	30

X_{L}

図 3.1. 行和 $(12, 18)$ と列和 $(5, 4, 3, 7, 5, 6)$ の与えられた分割表集合に対する X_{U} と X_{L}。

アルゴリズム 2

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。

定理 3.1 アルゴリズム 2 は確率 1 で停止して、分割表を 1 つ出力する。その分割表は Ξ 上の一様分布に従う確率変数の実現値である。

定理 3.1 はアルゴリズム 2 がパーフェクトサンプリング法であることを保証する。次の小節では、提案したマルコフ連鎖の単調性を示すことで、定理 3.1 を示す。

3.2 マルコフ連鎖の単調性

2.1 節で 2 つの定理を紹介した。したがって、 定理 3.1 を示すにはアルゴリズム 2 が単調 CFTP であることを示せばよい。そのために、この小節では Ξ 上に半順序を導入し、X_{U} と X_{L} が唯一の最大元および最小元の対であることと、マルコフ連鎖 $\mathcal{M}_{\mathrm{C}}$ が単調であることを示す。

任意の $X \in \Xi$ に対して、累積和ベクトル (*cumulative sum vector*) $f_X \in \mathbb{Z}_+^{n+1}$ を

$$f_X(i) \stackrel{\text{def.}}{=} \begin{cases} 0 & (i = 0), \\ X[1,1] + \cdots + X[1,i] & (i \in \{1, \ldots, n\}), \end{cases}$$

と定義する。ただし、$f_X \stackrel{\text{def.}}{=} (f_X(0), f_X(1), \ldots, f_X(n))$ である。この定義から明らかに、Ξ と $\{f_X \mid X \in \Xi\}$ の間には全単射が存在する。任意の対 $X, Y \in \Xi$ に対して、$X \succeq Y$ の必要十分条件を $f_X - f_Y \geq \mathbf{0}$ とする。2 項関係 "$\succeq$" が Ξ 上の半順序であることは明らかである。

まず、この半順序集合における最大元と最小元について考える。

補題 3.2 与えられた半順序集合 $(\Xi, \succeq)$ において、任意の $X \in \Xi$ に対して、$X_{\mathrm{U}} \succeq X \succeq X_{\mathrm{L}}$ が成り立つ。

証明: 明らかに、$\forall X \in \Xi,\, \forall i \in \{0,1,\ldots,n\},\, 0 \le f_X(i) \le r_1$ が成り立つ。まず、任意の $X \in \Xi$ に対して $f_{X_{\rm U}}(i) \ge f_X(i)$ を i に関する帰納法で示す。$i=0$ に対して $f_{X_{\rm U}}(0) = f_X(0) = 0$ が成り立つ。いま $f_{X_{\rm U}}(i-1) \ge f_X(i-1)$ と仮定する。この時、$f_{X_{\rm U}}(i) = f_{X_{\rm U}}(i-1) + X_{\rm U}[1,i] = \min\{f_{X_{\rm U}}(i-1)+s_i, r_1\} \ge \min\{f_X(i-1)+s_i, r_1\} \ge f_X(i)$ が成り立つ。したがって任意の $X \in \Xi$ に対して $X_{\rm U} \succeq X$ を得る。

次に、任意の $X \in \Xi$ に対して $f_X(i) \ge f_{X_{\rm L}}(i)$ を i に関する帰納法で示す。$i=n$ に対して $f_X(n) = f_{X_{\rm L}}(n) = r_1$ が成り立つ。いま $f_X(i) \ge f_{X_{\rm L}}(i)$ と仮定する。この時、$f_{X_{\rm L}}(i-1) = f_{X_{\rm L}}(i) - X_{\rm L}[1,i] = \max\{f_{X_{\rm L}}(i) - s_i, 0\} \le \max\{f_X(i) - s_i, 0\} \le f_X(i-1)$ が成り立つ。したがって任意の $X \in \Xi$ に対して $X \succeq X_{\rm L}$ を得る。 □

いま $X, Y \in \Xi$ に対して、ある k が存在して

$$f_X(i) - f_Y(i) = \begin{cases} 1 & (i=k), \\ 0 & (\text{otherwise}). \end{cases}$$

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。すなわち $X \cdot\!\succ_k Y$ の必要十分条件は

$$X[1,i] - Y[1,i] = \begin{cases} +1 & (i=k), \\ -1 & (i=k+1), \\ 0 & (\text{otherwise}). \end{cases}$$

である。

補題 3.3 相異なる 2 つの状態 $X, Y \in \Xi$ に対して $X \succeq Y$ が成り立つとき、$\exists Z \in \Xi,\, X \cdot\!\succ Z \succeq Y$ が成り立つ。

証明: 次の 2 つの場合を考える必要がある。

1. まず、$\exists k' \in \{0,1,\ldots,n-1\},\, f_X(k') > f_Y(k')$ かつ $X[1,k'+1] < s_{k'+1}$ の場合を考える。
   添え字 k は上の条件を満たす最小のものとする。この時 $f_X(0) = f_Y(0) = 0$ より明らかに $k \ge 1$ である。ここで $X[1,k] > 0$ を示す。もし $f_X(k-1) > f_Y(k-1)$ ならば、k の最小性から $X[1,k] = s_k > 0$ である。$f_X(k-1) \le f_Y(k-1)$ の時、$X \succeq Y$ から $f_X(k-1) = f_Y(k-1)$ であり、$X[1,k] = f_X(k) - f_X(k-1) > f_Y(k) - f_Y(k-1) = Y[1,k] \ge 0$ である。したがって分割表 Z を

   $$\begin{aligned} Z[1,l] &= \begin{cases} X[1,k]-1 & (l=k), \\ X[1,k+1]+1 & (l=k+1), \\ X[1,l] & (\text{otherwise}), \end{cases} \\ Z[2,l] &= s_l - Z[1,l] \quad (l=1,\ldots,n), \end{aligned}$$

   のように定めると Z は $Z \in \Xi$ かつ $X \cdot\!\succ Z \succeq Y$ である。

2. 次に、$\forall k' \in \{0,1,\ldots,n-1\},\, f_X(k') > f_Y(k') \Rightarrow X[1,k'+1] = s_{k'+1}$ の場合を考える。
   以下では$\forall k \in \{0,\ldots,n-1\},\, X[1,k+1] \ge Y[1,k+1]$ を示す。もし $f_X(k) > f_Y(k)$ ならば、場合 2 の仮定より $X[1,k+1] = s_{k+1} \ge Y[1,k+1]$ である。このとき $f_X(k) \le f_Y(k)$ ならば、$X \succeq Y$ より、$f_X(k) = f_Y(k)$ かつ $f_X(k+1) \ge f_Y(k+1)$ である。したがって $X[1,k+1] = f_X(k+1) - f_X(k) \ge f_Y(k+1) - f_Y(k) = Y[1,k+1]$ である。いま $X[1,1] + \cdots + X[1,n] = Y[1,1] + \cdots + Y[1,n]$ より、$\forall k \in \{0,1,\ldots,n-1\},\, X[1,k+1] \ge Y[1,k+1]$ から $X = Y$ となり矛盾である。 □

次の補題はマルコフ連鎖の単調性を示す上で鍵となるものである。

補題 3.4 もし 2 つの相異なる状態 $X, Y \in \Xi$ 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。任意の添え字 $i \neq \lfloor \lambda \rfloor$ に対しては $f_{X'}(i) = f_X(i)$ および $f_{Y'}(i) = f_Y(i)$ が明らかに成り立つので定理の仮定 $X \succeq Y$ より $f_{X'}(i) - f_{Y'}(i) = f_X(i) - f_Y(i) \geq 0$ が成り立つ。そこで $i = \lfloor \lambda \rfloor$ の場合を考えると、

$$\begin{aligned} f_{X'}(\lfloor \lambda \rfloor) - f_{Y'}(\lfloor \lambda \rfloor) &= (f_{X'}(\lfloor \lambda \rfloor - 1) + X'[1, \lfloor \lambda \rfloor]) - (f_{Y'}(\lfloor \lambda \rfloor - 1) + Y'[1, \lfloor \lambda \rfloor]) \\ &= \{f_X(\lfloor \lambda \rfloor - 1) - f_Y(\lfloor \lambda \rfloor - 1)\} + (X'[1, \lfloor \lambda \rfloor] - Y'[1, \lfloor \lambda \rfloor]) \\ &= \{f_X(\lfloor \lambda \rfloor - 1) - f_Y(\lfloor \lambda \rfloor - 1)\} \\ &\quad + \min\{a_X, s_{\lfloor \lambda \rfloor}\} - \lfloor (\lambda - \lfloor \lambda \rfloor)\, \theta_X \rfloor - \min\{a_Y, s_{\lfloor \lambda \rfloor}\} + \lfloor (\lambda - \lfloor \lambda \rfloor)\, \theta_Y \rfloor. \\ &= \begin{cases} \Delta\eta + \Delta\theta & (\lfloor \lambda \rfloor \neq k+1), \\ 1 + \Delta\eta + \Delta\theta & (\lfloor \lambda \rfloor = k+1), \end{cases} \end{aligned}$$

である。ここで $a_X \stackrel{\text{def.}}{=} a_X(\lfloor \lambda \rfloor)$, $a_Y \stackrel{\text{def.}}{=} a_Y(\lfloor \lambda \rfloor)$, $\theta_X \stackrel{\text{def.}}{=} \theta_X(\lfloor \lambda \rfloor)$, $\theta_Y \stackrel{\text{def.}}{=} \theta_Y(\lfloor \lambda \rfloor)$ であり ((3.1), (3.3) 参照) $\Delta\eta \stackrel{\text{def.}}{=} \min\{a_X, s_{\lfloor \lambda \rfloor}\} - \min\{a_Y, s_{\lfloor \lambda \rfloor}\}$, $\Delta\theta \stackrel{\text{def.}}{=} -\lfloor (\lambda - \lfloor \lambda \rfloor)\, \theta_X \rfloor + \lfloor (\lambda - \lfloor \lambda \rfloor)\, \theta_Y \rfloor$ とする。

1. $\lfloor \lambda \rfloor = k - 1$ の場合を考える。このとき $a_X = a_Y + 1$ かつ $b_X = b_Y - 1$ である。ただし、$b_X \stackrel{\text{def.}}{=} b_X(\lfloor \lambda \rfloor)$, $b_Y \stackrel{\text{def.}}{=} b_Y(\lfloor \lambda \rfloor)$ である ((3.2) 参照)。
   (a) もし $a_Y \geq s_{\lfloor \lambda \rfloor}$ ならば $\Delta\eta = 0$ かつ $\theta_X \leq \theta_Y$ が成り立つ。したがって $\Delta\theta \geq 0$ となり、$f_{X'}(\lfloor \lambda \rfloor) - f_{Y'}(\lfloor \lambda \rfloor) \geq 0$ が成り立つ。
   (b) もし $a_Y < s_{\lfloor \lambda \rfloor}$ ならば $\Delta\eta = 1$ and $\theta_X \leq \theta_Y + 1$ が成り立つ。したがって $\Delta\theta \geq -1$ となり、$f_{X'}(\lfloor \lambda \rfloor) - f_{Y'}(\lfloor \lambda \rfloor) \geq 0$ が成り立つ。
2. $\lfloor \lambda \rfloor = k + 1$ の場合を考える。このとき $a_X = a_Y - 1$ かつ $b_X = b_Y + 1$ である。
   (a) もし $a_X \geq s_{\lfloor \lambda \rfloor}$ ならば $1 + \Delta\eta \geq 1$ かつ $\theta_X \leq \theta_Y + 1$ が成り立つ。したがって $\Delta\theta \geq -1$ となり、$f_{X'}(\lfloor \lambda \rfloor) - f_{Y'}(\lfloor \lambda \rfloor) \geq 0$ が成り立つ。
   (b) もし $a_X < s_{\lfloor \lambda \rfloor}$ ならば $1 + \Delta\eta \geq 0$ が成り立つ。いま $a_X + b_X = a_Y + b_Y = s_{\lfloor \lambda \rfloor} + s_{\lfloor \lambda \rfloor + 1}$ かつ $\theta_X \leq \theta_Y$ に注意すると、$\Delta\theta \geq 0$ であり、$f_{X'}(\lfloor \lambda \rfloor) - f_{Y'}(\lfloor \lambda \rfloor) \geq 0$ が成り立つ。
3. それ以外の場合、すなわち $\lfloor \lambda \rfloor \neq k + 1$ かつ $\lfloor \lambda \rfloor \neq k - 1$ の時、$a_X = a_Y$, $\Delta\eta = 0$, $\Delta\theta = 0$ となり、$f_{X'}(\lfloor \lambda \rfloor) - f_{Y'}(\lfloor \lambda \rfloor) = 0$ が成り立つ。

以上より $f_{X'} \geq f_{Y'}$ が成り立ち、したがって $\phi(X, \lambda) \succeq \phi(Y, \lambda)$ を得る。 □

補題 3.5 マルコフ連鎖 $\mathcal{M}_{\mathrm{C}}$ は単調である。すなわち $\forall \lambda \in [1, n)$, $\forall X, \forall Y \in \Xi$, $X \succeq Y \Rightarrow \phi(X, \lambda) \succeq \phi(Y, \lambda)$ が成り立つ。

証明: 相異なる 2 つの状態 $X, Y \in \Xi$ が $X \succeq Y$ を満たす時、補題 3.3 を繰り返し適用することで、分割表の列 $X = Z_0, Z_1, \ldots, Z_R = Y$ が存在して $Z_i \in \Xi$ $(0 \leq i \leq R)$ かつ $Z_0 \cdot\succ Z_1 \cdot\succ \cdots \cdot\succ Z_R$ となることが示される。したがって、補題 3.4 より任意の $\lambda \in [1, n)$ に対して $\phi(Z_0, \lambda) \succeq \phi(Z_1, \lambda) \succeq \cdots \succeq \phi(Z_R, \lambda)$ が成り立つ。よって $\forall \lambda \in [1, n)$, $\phi(X, \lambda) \succeq \phi(Y, \lambda)$ を得る。 □

最後に、提案したアルゴリズムがパーフェクトサンプリング法であることを示す。

定理 3.1 の証明: 補題 3.5 よりマルコフ連鎖 $\mathcal{M}_{\mathrm{C}}$ は単調であり、補題 3.2 より X_{U} と X_{L} はそれぞれ唯一の最大および最小元である。したがってアルゴリズム 2 は単調 CFTP であり、定理 2.3 および定理 2.5 から定理 3.1 を得る。 □

3.3 アルゴリズムの計算時間

以下、アルゴリズムの計算時間について議論する。この節では次の条件を仮定する。この条件は計算時間の多項式性を保障する上で本質的である。

条件 **3.1** 列和ベクトル $\boldsymbol{s}$ は $s_1 \geq s_2 \geq \cdots \geq s_n$ を満たす。

次の定理は本論文の主要な結果のひとつである。

定理 **3.6** 条件 3.1 の下、アルゴリズム 2 の計算時間の期待値は $\mathrm{O}(n^3 \ln N)$ である。ただし、n は列数であり、N は Ξ に属する各分割表の総計値とする。

以下では、coalescence time $T_* \in \mathbb{Z}_{++}$ を算定することで定理 3.6 を示す。ただし、coalescence time は $T_* \stackrel{\text{def.}}{=} \min\{t > 0 \mid \exists y \in \Omega,\ \forall x \in \Omega,\ y = \Phi^0_{-t}(x, \boldsymbol{\Lambda})\}$ と定義される確率変数である。

2.3 節で述べたように coalescence time と mixing time の間には関係があることが知られている。まず、提案したマルコフ連鎖 $\mathcal{M}_{\mathrm{C}}$ の mixing rate を path coupling 定理を用いて算定する。この補題において条件 3.1 は重要である。

補題 **3.7** 条件 3.1 の下で提案したマルコフ連鎖 $\mathcal{M}_{\mathrm{C}}$ の mixing rate τ は $\tau \leq n(n-1)^2(1 + \ln(nN))$ を満たす。

証明: 単純無向グラフ $G = (\Xi, \mathcal{E})$ は頂点集合 Ξ と次に定義する枝集合 $\mathcal{E}$ をもつ。任意の頂点対 $\{X, Y\}$ が $\mathcal{E}$ に属する必要十分条件は $(1/2)\sum_{j=1}^{n} |X[1,j] - Y[1,j]| = 1$ 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 (*supporting 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 2 つの分割表の対 $X, Y \in \Xi$ 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。明らかに G の直径、すなわち $\max\{d(X,Y)\}$ は nN 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 3.1 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、

$$\mathrm{E}[d(X',Y')] \leq \beta d(X,Y),\ \beta = 1 - 1/(n(n-1)^2), \tag{3.4}$$

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 $X[1,j_2] = Y[1,j_2] - 1$ を仮定する。以下の証明に現れる $a_X, a_Y, b_X, b_Y, \theta_X, \theta_Y$ の定義は、補題 3.4 の証明で使ったものと同じである。

1. $\lfloor\Lambda\rfloor = j_2 - 1$ の場合について、$\mathrm{E}[d(X',Y') \mid \lfloor\Lambda\rfloor = j_2 - 1] \leq d(X,Y) - (1/2)(n - j_2 + 1)/(n-1)$ を示す。
   (a) $j_1 = j_2 - 1$ の時、条件 3.1 より $X' = Y'$ である。ゆえに $d(X',Y') = 0$。
   (b) $j_1 \neq j_2 - 1$ かつ $\theta_X = \theta_Y$ の時。条件 3.1 より $s_{j_2-1} \geq s_{j_2}$ である。いま、$\theta_X = \theta_Y$ かつ $a_Y > a_X$ かつ $b_Y < b_X$ なので $\theta_X = \theta_Y = \min\{s_{j_2-1}, s_{j_2}\} = s_{j_2}$ となる。したがって $s_{j_2-1} \geq a_Y > a_X \geq s_{j_2}$ となり関数 ϕ の定義より $X'[1, j_2-1] = a_X - (\Lambda - \lfloor\Lambda\rfloor)\theta_X$ と $Y'[1, j_2-1] = a_Y - (\Lambda - \lfloor\Lambda\rfloor)\theta_Y$ を得る。この時 $a_X = a_Y - 1$ から $X'[1, j_2-1] = Y'[1, j_2-1] - 1$ である。また、$X'[1,j_2] = a_X - X'[1, j_2-1]$ かつ $Y'[1,j_2] = a_Y - Y'[1, j_2-1]$ より $X'[1,j_2] = Y'[1,j_2]$ である。ゆえにこの条件のもと確率 1 で $d(X',Y') = d(X,Y) - (n - j_2 + 1)/(n-1)$ である。

(c) $j_1 \neq j_2 - 1$ かつ $\theta_X \neq \theta_Y$ の時。明らかに、$|\theta_X - \theta_Y| = 1$ である。まず、$\theta_X = \theta_Y - 1$ の場合を考える。この時、2 つの場合を考えれば良い。$\lfloor(\Lambda - \lfloor\Lambda\rfloor)\theta_X\rfloor = \lfloor(\Lambda - \lfloor\Lambda\rfloor)\theta_Y\rfloor$ と $\lfloor(\Lambda - \lfloor\Lambda\rfloor)\theta_X\rfloor = \lfloor(\Lambda - \lfloor\Lambda\rfloor)\theta_Y\rfloor - 1$ の 2 つの場合を考えれば十分である。前者の場合、$X'[1, j_2 - 1] = Y'[1, j_2 - 1]$ かつ $X'[1, j_2] = Y'[1, j_2] - 1$ なので、$d(X', Y') = d(X, Y)$ である。後者の場合、$X'[1, j_2 - 1] = Y'[1, j_2 - 1] - 1$ かつ $X'[1, j_2] = Y'[1, j_2]$ なので、$d(X', Y') = d(X, Y) - (n - j_2 + 1)/(n - 1)$ である。2 つの場合は確率 $1/2$ で現れるので、$\mathrm{E}[d(X', Y') \mid \lfloor\Lambda\rfloor = j_2 - 1, j_1 \neq j_2 - 1, \theta_X = \theta_Y - 1] = d(X, Y) - (1/2)(n - j_2 + 1)/(n - 1)$ である。残った $\theta_X = \theta_Y + 1$ の場合についても同様の証明が行える。

2. $\lfloor\Lambda\rfloor = j_2$ の場合について、$\mathrm{E}[d(X', Y') \mid \lfloor\Lambda\rfloor = j_2] \leq d(X, Y) + (1/2)(n - j_2)/(n - 1)$ を示す。

   (a) $\theta_X = \theta_Y$ の時、1-(b) の場合と同様に $X'[1, j_2] = Y'[1, j_2] - 1$ and $X'[1, j_2 + 1] = Y'[1, j_2 + 1]$ を得る。ゆえにこの条件の下確率 1 で $d(X', Y') = d(X, Y)$ である。

   (b) $\theta_X \neq \theta_Y$ の時、1-(c) の場合と同様に条件付確率 $1/2$ で $d(X', Y') = d(X, Y)$ と条件付確率 $1/2$ で $d(X', Y') = d(X, Y) + (n - j_2)/(n - 1)$ を示すことができる。ゆえにこの場合 $\mathrm{E}[d(X', Y') \mid \lfloor\Lambda\rfloor = j_2, \theta_X \neq \theta_Y] = d(X, Y) + (1/2)(n - j_2)/(n - 1)$ となる。

3. $\lfloor\Lambda\rfloor \neq j_2 - 1$ かつ $\lfloor\Lambda\rfloor \neq j_2$ の場合。この時、$\{X', Y'\}$ もまた G の枝となる。いま、$\{j_1', j_2'\}$ を $\{X', Y'\}$ の支持対とすると、明らかに $j_2 = \max\{j_1', j_2'\}$ である。したがって $d(X', Y') = d(X, Y)$ を得る。

それぞれの場合の生起する確率は Case 1 が $1/(n-1)$、Case 2 が $1/(n-1)$ 以下、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。グラフ G の直径が nN で押さえられることから、定理 2.2 より、mixing rate τ は

$$\tau \leq n(n-1)^2(1 + \ln(nN))$$

を満たす。 □

次に、coalescence time を算定する。2.3 節の結論をそのまま用いて coalescence time を算定することもできるが、より良い上界を与えるために 2.3 節の証明と同様の技法を用いて直接算定する。

補題 3.8 条件 3.1 の下で、マルコフ連鎖 $\mathcal{M}_{\mathrm{C}}$ の coalescence time T_* は $\mathrm{E}[T_*] = \mathrm{O}(n^3 \ln N)$ を満たす。

証明: 単純無向グラフ $G = (\Xi, \mathcal{E})$ と G 上の任意の頂点対 $X, Y \in \Xi$ に対する距離 $d(X, Y)$ は補題 3.7 で定義されたものとする。また、$D_G \stackrel{\text{def.}}{=} d(X_{\mathrm{U}}, X_{\mathrm{L}})$ と $\tau_0 \stackrel{\text{def.}}{=} n(n-1)^2(1 + \ln D_G)$ を定義する。補題 3.7 で得られた不等式 (3.4) を用いると、

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

が得られる。ここで coalescence time の持つ submultiplicativity ([40]) 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

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

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

最後にアルゴリズム 2 の計算時間について議論する。

定理 3.6 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 2 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 2 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 2 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 $\mathrm{O}(\mathrm{E}[2T_*] + \mathrm{E}[8T_*] + \mathrm{E}[K+1]n) = \mathrm{O}(\mathrm{E}[T_*]) = \mathrm{O}(n^3 \ln N)$ となる。 □

なお、条件 3.1 は $\mathrm{O}(n \ln n)$ で得られる。

第4章

離散化Dirichlet分布に従うパーフェクトサンプリング法

本章では、離散化 Dirichlet 分布に従う確率変数ベクトルのパーフェクトサンプリング法について述べる。

4.1 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 3

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。

定理 4.1 アルゴリズム 3 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。

上の定理によりアルゴリズム 3 はパーフェクトサンプリングアルゴリズムであることが保証される。次の小節ではマルコフ連鎖 $\mathcal{M}_{\mathrm{D}}$ の単調性を示すことで上記定理を示す。

4.2 マルコフ連鎖の単調性

2.1 節で 2 つの定理を紹介した。したがって、 定理 4.1 を示すにはアルゴリズム 3 が単調 CFTP であることを示せばよい。そのために、この小節では Ω 上に半順序を導入し、X_{U} と X_{L} が唯一の最大元および最小元の対であることと、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 $\mathrm{c}_X \in \mathbb{Z}_+^{n+1}$

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

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。

次に、単調性を証明する鍵となる補題を示す。

補題 4.2 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、$b' \stackrel{\text{def.}}{=} X_{\lfloor \lambda \rfloor} + X_{\lfloor \lambda \rfloor + 1}$ 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、$b'' \stackrel{\text{def.}}{=} Y_{\lfloor \lambda \rfloor} + Y_{\lfloor \lambda \rfloor + 1}$ 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、この後で示す補題 4.3 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

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

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 (X', Y') は

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

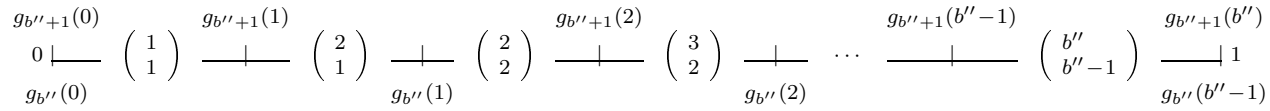


図 4.1. alternating inequalities の図。図中で $g_{b''}(k)$, $g_{b''+1}(k)$ はそれぞれ $g_{b''}(k|u_{j-1}, u_j)$, $g_{b''+1}(k|u_{j-1}, u_j)$ を表す。

であることがわかる (図 4.1 参照)。以上のことから $X'_{j-1} \geq Y'_{j-1}$ 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、$1 + X'_{j+1} \geq Y'_{j+1}$ を示せば十分である。

補題 4.3 から、Case 2 と同様に次の alternating inequalities が得られる。

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

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 (図 4.2 参照)。以上のことから $1 + X'_{j+1} \geq Y'_{j+1}$ が得られた。 □

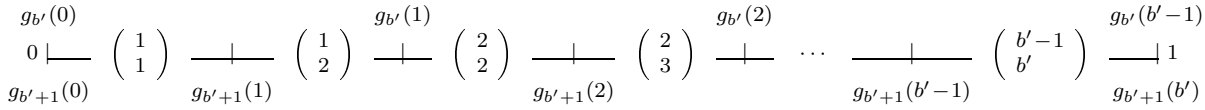


図 4.2. alternating inequalities の図。図中で $g_{b'}(k)$, $g_{b'+1}(k)$ はそれぞれ $g_{b'}(k|u_{j+1}, u_{j+2})$, $g_{b'+1}(k|u_{j+1}, u_{j+2})$ を表す。

補題 4.3

$$\begin{aligned} &\forall b \in \{2, 3, \ldots\},\ \forall u_i, \forall u_j \geq 0,\ \forall k \in \{1, 2, \ldots, b\}, \\ &g_{b+1}(k-1|u_i, u_j) \leq g_b(k-1|u_i, u_j) \leq g_{b+1}(k|u_i, u_j). \end{aligned}$$

証明: 以下では命題の 2 番目の不等式を示す。1 番目の不等式は同様に示すことができる。

簡単のため $C_{b+1} = C(u_i, u_j, b+1)$, $C_b = C(u_i, u_j, b)$ と表す。いま $g_b(k|u_i, u_j)$ の定義から

$$\begin{aligned} H(k) &\overset{\text{def.}}{=} g_{b+1}(k|u_i, u_j) - g_b(k-1|u_i, u_j) \\ &= \textstyle\sum_{l=1}^{k} C_{b+1} l^{u_i - 1}(b-l+1)^{u_j - 1} - \sum_{l=1}^{k-1} C_b l^{u_i - 1}(b-l)^{u_j - 1} \end{aligned}$$

$$
\begin{aligned}
&= (1 - C_{b+1}\textstyle\sum_{l=k+1}^{b} l^{u_i-1}(b-l+1)^{u_j-1}) - (1 - C_b \textstyle\sum_{l=k}^{b-1} l^{u_i-1}(b-l)^{u_j-1}) \\
&= C_b \textstyle\sum_{l=k+1}^{b} (l-1)^{u_i-1}(b-l+1)^{u_j-1} - C_{b+1}\textstyle\sum_{l=k+1}^{b} l^{u_i-1}(b-l+1)^{u_j-1} \\
&= \textstyle\sum_{l=k+1}^{b} (C_b(l-1)^{u_i-1}(b-l+1)^{u_j-1} - C_{b+1} l^{u_i-1}(b-l+1)^{u_j-1}) \\
&= \textstyle\sum_{l=k+1}^{b} C_b l^{u_i-1}(b-l+1)^{u_j-1} \left(\left(1-\frac{1}{l}\right)^{u_i-1} - \frac{C_{b+1}}{C_b} \right)
\end{aligned}
$$

を得る。同様に

$$
\begin{aligned}
H(k) &= C_{b+1}\textstyle\sum_{l=1}^{k} l^{u_i-1}(b-l+1)^{u_j-1} - C_b \textstyle\sum_{l=1}^{k-1} l^{u_i-1}(b-l)^{u_j-1} \\
&\geq C_{b+1}\textstyle\sum_{l=2}^{k} l^{u_i-1}(b-l+1)^{u_j-1} - C_b \textstyle\sum_{l=2}^{k} (l-1)^{u_i-1}(b-l+1)^{u_j-1} \\
&= \textstyle\sum_{l=2}^{k} (C_{b+1} l^{u_i-1}(b-l+1)^{u_j-1} - C_b (l-1)^{u_i-1}(b-l+1)^{u_j-1}) \\
&= \textstyle\sum_{l=2}^{k} C_b l^{u_i-1}(b-l+1)^{u_j-1} \left(\frac{C_{b+1}}{C_b} - \left(1-\frac{1}{l}\right)^{u_i-1} \right)
\end{aligned}
$$

も示される。関数 $h : \{2,3,\ldots,b\} \to \mathbb{R}$ を導入し、$h(l) \stackrel{\text{def.}}{=} \left(1-\frac{1}{l}\right)^{u_i-1} - \frac{C_{b+1}}{C_b}$ と定義すると、次の不等式が得られる。

$$
\begin{aligned}
H(k) &= \textstyle\sum_{l=k+1}^{b} C_b l^{u_i-1}(b-l+1)^{u_j-1} h(l) && (4.1) \\
&\geq -\textstyle\sum_{l=2}^{k} C_b l^{u_i-1}(b-l+1)^{u_j-1} h(l). && (4.2)
\end{aligned}
$$

(a) $u_i \geq 1$ の場合を考える。

いま $u_i - 1 \geq 0$ なので $h(l)$ は単調非減少関数である。したがって $h(k) \geq 0$ が成り立てば $0 \leq h(k) \leq h(k+1) \leq \cdots \leq h(b)$ となり、等式 (4.1) から $H(k) \geq 0$ が言える。もし $h(k) < 0$ なら $h(2) \leq h(3) \leq \cdots \leq h(k) < 0$ となるので不等式 (4.2) を用いて $H(k) \geq -\sum_{l=2}^{k} C_b l^{u_i-1}(b-l+1)^{u_j-1} h(l) \geq 0$ が得られる。

(b) $0 \leq u_i \leq 1$ の場合を考える。

この場合 $u_i - 1 \leq 0$ なので $h(l)$ は非増加関数である。もし不等式 $h(b) \geq 0$ が成り立てば、$h(2) \geq h(3) \geq \cdots \geq h(b) \geq 0$ となり等式 (4.1) から $H(k) \geq 0$ を得る。したがって $h(b) = (\frac{b-1}{b})^{u_i-1} - \frac{C_{b+1}}{C_b} \geq 0$ を示せば十分である。

以下この証明において簡単のため $\alpha_{i'} = u_{i'} - 1$ $(\forall i')$ で表す。関数 $H_0(b, \alpha_i, \alpha_j)$ を $H_0(b, \alpha_i, \alpha_j) \stackrel{\text{def.}}{=} (b-1)^{\alpha_i} C_{b+1}^{-1} - b^{\alpha_i} C_b^{-1}$ と定義する。もし条件

$$
-1 \leq \forall \alpha_i \leq 0,\ -1 \leq \forall \alpha_j,\ \forall b \in \{2,3,4,\ldots\},\ H_0(b, \alpha_i, \alpha_j) \geq 0
$$

がを満たせば、目標の各 $b \in \{2,3,4,\ldots\}$ に対して $h(b) \geq 0$ を得る。ここで $H_0(b, \alpha_i, \alpha_j)$ を式変形する。

$$
\begin{aligned}
H_0(b, \alpha_i, \alpha_j) &= (b-1)^{\alpha_i} \sum_{k=1}^{b} k^{\alpha_i}(b-k+1)^{\alpha_j} - b^{\alpha_i} \sum_{k=1}^{b-1} k^{\alpha_i}(b-k)^{\alpha_j} \\
&= \textstyle\sum_{k=1}^{b} (b-1)^{\alpha_i} k^{\alpha_i} (b-k+1)^{\alpha_j} \frac{(b-k)+(k-1)}{b-1} - b^{\alpha_i} \sum_{k=1}^{b-1} k^{\alpha_i}(b-k)^{\alpha_j} \\
&= \textstyle\sum_{k=1}^{b-1} \Big[(b-1)^{\alpha_i} k^{\alpha_i} (b-k+1)^{\alpha_j} \left(\frac{b-k}{b-1}\right) + (b-1)^{\alpha_i} (k+1)^{\alpha_i} (b-k)^{\alpha_j} \left(\frac{k}{b-1}\right) \\
&\qquad\qquad - b^{\alpha_i} k^{\alpha_i} (b-k)^{\alpha_j} \Big] \\
&= \textstyle\sum_{k=1}^{b-1} \frac{(b-1)^{\alpha_i} k^{\alpha_i} (b-k)^{\alpha_j}}{b-1} \left[\left(1+\frac{1}{b-k}\right)^{\alpha_j} (b-k) + \left(1+\frac{1}{k}\right)^{\alpha_i} k - \left(\frac{b}{b-1}\right)^{\alpha_i} (b-1) \right].
\end{aligned}
$$

以上から、関数

$$
H_1(b, \alpha_i, \alpha_j, k) \stackrel{\text{def.}}{=} \left(1+\frac{1}{b-k}\right)^{\alpha_j} (b-k) + \left(1+\frac{1}{k}\right)^{\alpha_i} k - \left(\frac{b}{b-1}\right)^{\alpha_i} (b-1)
$$

が任意の $k \in \{1,2,\ldots,b-1\}$ に対して非負であることを示せば十分である。いま $1 + 1/(b-k) > 1$ かつ $\alpha_j \geq -1$ より、

$$
H_1(b, \alpha_i, \alpha_j, k) \geq H_1(b, \alpha_i, -1, k) = \frac{(b-k)^2}{b-k+1} + \left(1+\frac{1}{k}\right)^{\alpha_i} k - \left(\frac{b}{b-1}\right)^{\alpha_i} (b-1).
$$

である。ここで関数 H_1 を α_i で微分すると、

$$\begin{aligned}\frac{\partial}{\partial\alpha_i}H_1(b,\alpha_i,-1,k) &= \left(1+\tfrac{1}{k}\right)^{\alpha_i} k\log\left(1+\tfrac{1}{k}\right) - \left(\tfrac{b}{b-1}\right)^{\alpha_i}(b-1)\log\left(\tfrac{b}{b-1}\right)\\ &= \left(1+\tfrac{1}{k}\right)^{\alpha_i}\log\left(1+\tfrac{1}{k}\right)^k - \left(1+\tfrac{1}{b-1}\right)^{\alpha_i}\log\left(1+\tfrac{1}{b-1}\right)^{(b-1)}\end{aligned}$$

となる。正整数の対 k,b は $1\le k\le b-1$ を満たすので、α_i が非正であることを考慮して $0\le(1+1/k)^{\alpha_i}\le(1+1/(b-1))^{\alpha_i}$ と $0\le\log(1+1/k)^k\le\log(1+1/(b-1))^{b-1}$ が得られる。したがって関数 $H_1(b,\alpha_i,-1,k)$ は $\alpha_i\le 0$ に関して単調非減少である。ゆえに

$$\begin{aligned}H_1(b,\alpha_i,-1,k) &\ge H_1(b,0,-1,k) = \tfrac{(b-k)^2}{b-k+1} + \left(1+\tfrac{1}{k}\right)^0 k - \left(\tfrac{b}{b-1}\right)^0(b-1)\\ &= \tfrac{(b-k)^2}{b-k+1} + k - b + 1 = \tfrac{(b-k)^2+1^2-(b-k)^2}{b-k+1} = \tfrac{1}{b-k+1}\ge 0\end{aligned}$$

が成り立つ。 □

補題 4.4 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 4.2 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。 □

最後に、提案したアルゴリズムがパーフェクトサンプリング法であることを示す。

定理 4.1 の証明: 補題 4.4 よりマルコフ連鎖 $\mathcal{M}_{\mathrm{D}}$ は単調であり、また X_{U} と X_{L} がそれぞれ唯一の最大および最小元であることは明らかである。したがってアルゴリズム 3 は単調 CFTP であり、定理 2.3 および定理 2.5 から定理 4.1 を得る。 □

4.3 アルゴリズムの計算時間

以下、アルゴリズムの計算時間について議論する。この節では次の条件を仮定する。

条件 4.1 Dirichlet パラメータは非増加順に並ぶ。すなわち $u_1\ge u_2\ge\cdots\ge u_n$ が成り立つ。

次の定理は本論分の主要な結果のひとつである。

定理 4.5 条件 4.1 の下、アルゴリズム 3 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。

以下では、coalescence time $T_*\in\mathbb{Z}_{++}$ を算定することで定理 4.5 を示す。ただし、coalescence time は $T_*\stackrel{\mathrm{def.}}{=}\min\{t>0\mid\exists y\in\Omega,\ \forall x\in\Omega,\ y=\Phi^0_{-t}(x,\mathbf{\Lambda})\}$ と定義される確率変数である。

前章と同様にまず、提案したマルコフ連鎖 $\mathcal{M}_{\mathrm{D}}$ の mixing rate を path coupling 定理を用いて算定する。この補題において条件 4.1 は重要である。

補題 4.6 条件 4.1 の下で提案したマルコフ連鎖 $\mathcal{M}_{\mathrm{D}}$ の mixing rate τ は $\tau\le n(n-1)^2(1+\ln n(\Delta-n)/2)$ 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 4.1 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、

$$\mathrm{E}[d(X',Y')]\leq\beta d(X,Y),\ \beta=1-1/(n(n-1)^2),\tag{4.3}$$

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 4.1 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 4.2 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。

ここで後述の補題 4.7 より、$u_{j_2-1}\geq u_{j_2}$ の場合

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

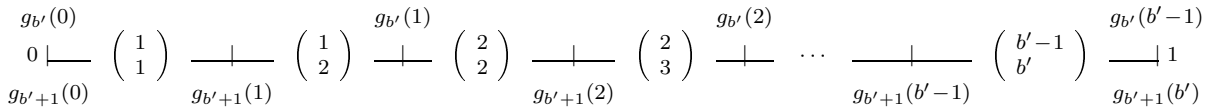


図 4.3. alternating inequalities の図。図中で $g_{b'}(k)$, $g_{b'+1}(k)$ はそれぞれ $g_{b'}(k|u_{j_2-1}, u_{j_2})$, $g_{b'+1}(k|u_{j_2-1}, u_{j_2})$ を表す。

が成り立つ (図 4.3 参照)。
条件 4.1 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。グラフ G の直径が $n(\Delta - n)/2$ で押さえられることから、定理 2.2 より、mixing rate τ は $\tau \leq n(n-1)^2(1 + \ln n(\Delta - n)/2)$ を満たす。 □

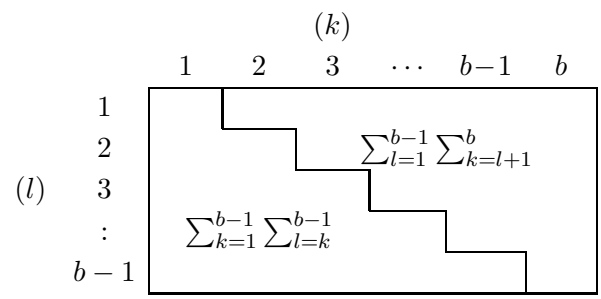
補題 4.7 $\forall b \in \{2, 3, \ldots\}$, $\forall u_i \geq \forall u_j$,

$$\sum_{k=1}^{b-1} [g_b(k|u_i, u_j) - g_{b+1}(k|u_i, u_j)] - \sum_{k=1}^{b-1} [g_{b+1}(k|u_i, u_j) - g_b(k-1|u_i, u_j)] \geq 0.$$

証明: 簡単のため $C_{b+1} = C(u_i, u_j, b+1)$ および $C_b = C(u_i, u_j, b)$ とし、次の等式変形を行う。

$$\begin{aligned} G &\stackrel{\text{def.}}{=} \textstyle\sum_{k=1}^{b-1} [g_b(k|u_i, u_j) - g_{b+1}(k|u_i, u_j)] - \sum_{k=1}^{b-1} [g_{b+1}(k|u_i, u_j) - g_b(k-1|u_i, u_j)] \\ &= \textstyle\sum_{k=1}^{b-1} g_b(k|u_i, u_j) - \sum_{k=1}^{b-1} g_{b+1}(k|u_i, u_j) \\ &\quad \textstyle - \sum_{k=1}^{b-1} g_{b+1}(k|u_i, u_j) + \sum_{k=1}^{b-1} g_b(k-1|u_i, u_j) \\ &= \textstyle\sum_{k=1}^{b-1} g_b(k|u_i, u_j) - \sum_{k=1}^{b-1} g_{b+1}(k|u_i, u_j) \\ &\quad \textstyle - \sum_{k=1}^{b-1} g_{b+1}(k|u_i, u_j) + \sum_{k=2}^{b-1} g_b(k-1|u_i, u_j) \\ &= \textstyle\sum_{k=1}^{b-1} C_b \sum_{l=1}^{k} l^{u_i - 1}(b-l)^{u_j - 1} - \sum_{k=1}^{b-1} C_{b+1} \sum_{l=1}^{k} l^{u_i - 1}(b-l+1)^{u_j - 1} \\ &\quad \textstyle - \sum_{k=1}^{b-1} C_{b+1} \sum_{l=1}^{k} l^{u_i - 1}(b-l+1)^{u_j - 1} + \sum_{k=2}^{b-1} C_b \sum_{l=1}^{k-1} l^{u_i - 1}(b-l)^{u_j - 1} \end{aligned}$$

$$
\begin{aligned}
&= C_b \textstyle\sum_{l=1}^{b-1}(b-l)l^{u_i-1}(b-l)^{u_j-1} - C_{b+1}\sum_{l=1}^{b-1}(b-l)l^{u_i-1}(b-l+1)^{u_j-1}\\
&\quad -C_{b+1}\textstyle\sum_{l=1}^{b-1}(b-l)l^{u_i-1}(b-l+1)^{u_j-1} + C_b\sum_{l=1}^{b-1}(b-l-1)l^{u_i-1}(b-l)^{u_j-1}\\
&= C_b \textstyle\sum_{l=1}^{b-1}(2b-2l-1)l^{u_i-1}(b-l)^{u_j-1} - C_{b+1}\sum_{l=1}^{b-1}(2b-2l)l^{u_i-1}(b-l+1)^{u_j-1}\\
&= C_bC_{b+1}\ \Big(\ C_{b+1}^{-1}\textstyle\sum_{l=1}^{b-1}(2b-2l-1)l^{u_i-1}(b-l)^{u_j-1}\\
&\qquad\qquad -C_b^{-1}\textstyle\sum_{l=1}^{b}(2b-2l)l^{u_i-1}(b-l+1)^{u_j-1}\Big)\\
&= C_bC_{b+1}\ \Big(\ \textstyle\sum_{k=1}^{b}k^{u_i-1}(b-k+1)^{u_j-1}\sum_{l=1}^{b-1}(2b-2l-1)l^{u_i-1}(b-l)^{u_j-1}\\
&\qquad\qquad -\textstyle\sum_{k=1}^{b-1}k^{u_i-1}(b-k)^{u_j-1}\sum_{l=1}^{b}(2b-2l)l^{u_i-1}(b-l+1)^{u_j-1}\Big)\\
&= C_bC_{b+1}\ \Big(\ \textstyle\sum_{k=1}^{b}\sum_{l=1}^{b-1}(2b-2l-1)(kl)^{u_i-1}((b-k+1)(b-l))^{u_j-1}\\
&\qquad\qquad -\textstyle\sum_{k=1}^{b-1}\sum_{l=1}^{b}(2b-2l)(kl)^{u_i-1}((b-k)(b-l+1))^{u_j-1}\Big)\\
&= C_bC_{b+1}\ \Big(\ \textstyle\sum_{k=1}^{b}\sum_{l=1}^{b-1}(2b-2l-1)(kl)^{u_i-1}((b-k+1)(b-l))^{u_j-1}\\
&\qquad\qquad -\textstyle\sum_{k=1}^{b}\sum_{l=1}^{b-1}(2b-2k)(kl)^{u_i-1}((b-l)(b-k+1))^{u_j-1}\Big)\\
&= C_bC_{b+1}\textstyle\sum_{k=1}^{b}\sum_{l=1}^{b-1}\left((2k-2l-1)(kl)^{u_i-1}((b-k+1)(b-l))^{u_j-1}\right)\\
&= \tfrac{C_bC_{b+1}}{2}\ \Big(\ \textstyle\sum_{k=1}^{b}\sum_{l=1}^{b-1}(2k-2l-1)(kl)^{u_i-1}((b-k+1)(b-l))^{u_j-1}\\
&\qquad\qquad +\textstyle\sum_{k=1}^{b}\sum_{l=1}^{b-1}(2k-2l-1)(kl)^{u_i-1}((b-k+1)(b-l))^{u_j-1}\Big)\\
&= \tfrac{C_bC_{b+1}}{2}\Big(\textstyle\sum_{k=1}^{b}\sum_{l=1}^{b-1}(2k-2l-1)(kl)^{u_i-1}((b-k+1)(b-l))^{u_j-1}\\
&\quad +\textstyle\sum_{k=1}^{b}\sum_{l=1}^{b-1}(2(b-k+1)-2(b-l)-1)\\
&\quad\ ((b-k+1)(b-l))^{u_i-1}((b-(b-k+1)+1)(b-(b-l)))^{u_j-1}\Big)\\
&= \tfrac{C_bC_{b+1}}{2}\ \Big(\ \textstyle\sum_{k=1}^{b}\sum_{l=1}^{b-1}(2k-2l-1)(kl)^{u_i-1}((b-k+1)(b-l))^{u_j-1}\\
&\qquad\qquad -\textstyle\sum_{k=1}^{b}\sum_{l=1}^{b-1}(2k-2l-1)((b-k+1)(b-l))^{u_i-1}(kl)^{u_j-1}\Big)
\end{aligned}
$$



$$
\begin{aligned}
&= \tfrac{C_bC_{b+1}}{2}\ \Big(\ \textstyle\sum_{l=1}^{b-1}\sum_{k=l+1}^{b}(2k-2l-1)(kl)^{u_i-1}((b-k+1)(b-l))^{u_j-1}\\
&\qquad\qquad +\textstyle\sum_{k=1}^{b-1}\sum_{l=k}^{b-1}(2k-2l-1)(kl)^{u_i-1}((b-k+1)(b-l))^{u_j-1}\\
&\qquad\qquad -\textstyle\sum_{l=1}^{b-1}\sum_{k=l+1}^{b}(2k-2l-1)((b-k+1)(b-l))^{u_i-1}(kl)^{u_j-1}\\
&\qquad\qquad -\textstyle\sum_{k=1}^{b-1}\sum_{l=k}^{b-1}(2k-2l-1)((b-k+1)(b-l))^{u_i-1}(kl)^{u_j-1}\Big)\\
&= \tfrac{C_bC_{b+1}}{2}\ \Big(\ \textstyle\sum_{l=1}^{b-1}\sum_{k=l+1}^{b}(2k-2l-1)(kl)^{u_i-1}((b-k+1)(b-l))^{u_j-1}\\
&\qquad\qquad -\textstyle\sum_{l=1}^{b-1}\sum_{k=l+1}^{b}(2k-2l-1)(l(k-1))^{u_i-1}((b-l+1)(b-k+1))^{u_j-1}\\
&\qquad\qquad -\textstyle\sum_{l=1}^{b-1}\sum_{k=l+1}^{b}(2k-2l-1)((b-k+1)(b-l))^{u_i-1}(kl)^{u_j-1}\\
&\qquad\qquad +\textstyle\sum_{l=1}^{b-1}\sum_{k=l+1}^{b}(2k-2l-1)((b-l+1)(b-k+1))^{u_i-1}(l(k-1))^{u_j-1}\Big)
\end{aligned}
$$

$$= \frac{C_b C_{b+1}}{2} \sum_{l=1}^{b-1} \sum_{k=l+1}^{b} (2k-2l-1) \Big((kl)^{u_i-1}((b-k+1)(b-l))^{u_j-1} - (l(k-1))^{u_i-1}((b-l+1)(b-k+1))^{u_j-1} - ((b-k+1)(b-l))^{u_i-1}(kl)^{u_j-1} + ((b-l+1)(b-k+1))^{u_i-1}(l(k-1))^{u_j-1} \Big).$$

いま関数 $G_0(k,l,u_i,u_j)$ を

$$G_0(k,l,u_i,u_j) \stackrel{\text{def.}}{=} \begin{pmatrix} (kl)^{u_i-1}((b-k+1)(b-l))^{u_j-1} \\ -(l(k-1))^{u_i-1}((b-l+1)(b-k+1))^{u_j-1} \\ -((b-k+1)(b-l))^{u_i-1}(kl)^{u_j-1} \\ +((b-l+1)(b-k+1))^{u_i-1}(l(k-1))^{u_j-1} \end{pmatrix}$$

と定義する。この時 $1 \le l < l+1 \le k \le b$ より、$(2k-2l-1) > 0$ は明らかに成り立つ。したがって $\forall l \in \{1,2,\ldots,b-1\}$, $\forall k \in \{2,3,\ldots,b\}$, $\forall u_i \ge \forall u_j$, $G_0(l,k,u_i,u_j) \ge 0$ を示せば十分である。明らかに

$$G_0(k,l,u_i,u_j) = \begin{pmatrix} (l(k-1))^{u_i-1}(b-k+1)^{u_j-1}\left((1+\frac{1}{k-1})^{u_i-1}(b-l)^{u_j-1} - (b-l+1)^{u_j-1}\right) \\ +((b-k+1)(b-l))^{u_i-1}l^{u_j-1}\left(-k^{u_j-1} + (1+\frac{1}{b-l})^{u_i-1}(k-1)^{u_j-1}\right) \end{pmatrix}$$

なので、$G_0(k,l,u_i,u_j)$ は u_i に関して非減少である。したがって $G_0(k,l,u_i,u_j) \ge G_0(k,l,u_j,u_j)$ を得る。ここで u_i に u_j を代入すると $G_0(k,l,u_i,u_j)$ の定義から $G_0(k,l,u_j,u_j) = 0$ となり、題意は示される。 □

次に coalescence time を算定する。

補題 4.8 条件 4.1 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 $G = (\Omega, \mathcal{E})$ と G 上の任意の頂点対 $X, Y \in \Omega$ に対する距離 $d(X,Y)$ は補題 4.6 で定義されたものとする。また $D \stackrel{\text{def.}}{=} d(X_{\mathrm{U}}, X_{\mathrm{L}})$ と $\tau_0 \stackrel{\text{def.}}{=} n(n-1)^2(1+\ln D)$ を定義する。補題 4.6 で得られた不等式 4.3 を用いると、

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

が得られる。ここで coalescence time の持つ submultiplicativity ([40]) を考慮すると、任意の $k \in \mathbb{Z}_+$ に対して、$\Pr(T_* > k\tau_0) \le (\Pr(T_* > \tau_0))^k \le (1/\mathrm{e})^k$ が得られる。したがって

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

となる。いま $n \le \Delta$ より、明らかに $D \le n(\Delta - n)/2 \le \Delta^2$ 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 3 の計算時間について議論する。

定理 4.5 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 3 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 3 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 3 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 4.1 は $\mathrm{O}(n \ln n)$ で得られる。

第5章

結論

本論文では周辺和の与えられた 2 行分割表を一様ランダムに生成する方法と離散化 Dirichlet 分布に従うサンプリング法について述べた。提案したサンプリング法はマルコフ連鎖を用いたサンプリング法で、それぞれの問題に対して rapidly mixing なマルコフ連鎖を提案した。さらに、これらのマルコフ連鎖が単調 (monotone) であることを示し、単調 CFTP (monotone CFTP) アルゴリズムを設計した。したがって CFTP 定理に基づき得られるサンプルは定常分布に厳密に従う。

単調 CFTP アルゴリズムを設計するには状態空間に半順序関係を導入する必要がある。本論文では累積和ベクトルを考慮することによって半順序関係を導入した。この累積和ベクトルを考慮し、Dyer and Greenhill のマルコフ連鎖の推移を制限することで、単調なマルコフ連鎖の設計を実現した。また、離散化 Dirichlet 分布に対しても、同様の手法をもって単調なマルコフ連鎖を設計した。

離散化 Dirichlet 分布ではマルコフ連鎖の単調性および rapidly mixing を証明するために alternating inequalty を導入した。本論文の結果はマルコフ連鎖の推移がこの不等式を満たすならば単調性および rapidly mixing の証明が今回と同様の方法で示せることを示唆する。

本論文ではアルゴリズムの計算時間の算定に mixing time の算定手法である Path Coupling を利用した。この手法を用いるためには状態空間に距離を導入する必要があるが、既存の研究と同様の距離を導入すると算定が甘く擬多項式時間となってしまうため、より精密な距離を考え弱多項式時間の上限を与えた。この点に関して計算機実験を行いこの算定がオーダーの意味でタイトであることを示唆する結果を得た。計算機実験の詳細は付録 B に記す。また、計算機実験では条件 3.1 および条件 4.1 が算定において本質的であることを示唆する結果についても得られた。

本論文で提案したアルゴリズムは素朴な CFTP アルゴリズムである。しかし、このアルゴリズムは乱数の列を記憶する必要があり、記憶容量の点で効率的でない。この点について、2000 年に Wilson の提案した Read Once アルゴリズムを適用することで解決できる [49]。Read Once アルゴリズムについては付録 A に記す。

提案したアルゴリズムではマルコフ連鎖の各推移が定数時間で行えることを仮定している。しかし、Dirichlet 分布では各推移を単位時間で行うことは困難である。サンプルひとつあたり coalescence time 回の推移が必要なことを考えれば乱数からベータ分布への変換をあらかじめ準備しておけばサンプルの個数が十分大きいときには amortize なアルゴリズムになる。

最後に、$m \times n$ 分割表の一様生成に関して、rapidly mixing なマルコフ連鎖が存在するか否か、単調なマルコフ連鎖が存在するか否か、が共に未解決問題として残されている。

謝辞

指導教官である数理第五研究室の松井知己助教授には、本研究だけでなく、修士課程における研究全般について温かくご指導頂きました。心より感謝いたします。数理第五研究室の杉原厚吉教授には、研究活動への多大なご支援をいただきましたことを厚くお礼申し上げます。数理第五研究室の西田徹志助手には研究環境について様々なご尽力をいただきました。また、数理第五研究室、数理第二研究室の皆様、他多くの方々に研究に対する助言を頂きました。すべての方々に感謝いたします。

参考文献

[1] A. Agresti, "A survey of exact inference for contingency tables," *Statistical Science*, **7** (1992), pp. 131–153.

[2] A. Agresti, *Categorical Data Analysis*, John Wiley & Sons, 2002.

[3] D. Aldous, "Random walks on finite groups and rapidly mixing Markov chains," in *Séminarie de Probabilitiés XVII 1981/1982*, vol. 986 of Springer-Verlag Lecture Notes in Mathematics, D. Dold and B. Eckmann, ed., Springer-Verlag, New York, 1983, pp. 243–297.

[4] S. Aoki and A. Takemura, "Minimal basis for connected Markov chain over $3\times3\times K$ contingency tables with fixed two-dimensional amrginals," *Austrarian and New Zealand Journal of Statistics*, **45** (2003), pp. 229–249.

[5] S. Boyd, P. Diaconis and L. Xiano, "Fastest Mixing Markov Chain on A Graph," submitted to *SIAM Review, problems and techniques section*, 2003, (available from http://www.stanford.edu/~boyd/fmmc.html)

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

[7] R. Bubley and M. Dyer, "Path coupling : a technique for proving rapid mixing in Markov chains," *Proceedings of the 38th Annual Symposium on Foundations of Computer Science* (FOCS 1997), pp. 223–231.

[8] T. L. Burr, "Quasi-equilibrium theory for the distribution of rare alleles in a subdivided population: justification and implications," *Theoretical Population Biology*, **57** (2000) 297–306.

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

[10] M. Cryan and M. Dyer, "A polynomial-time algorithm to approximately count contingency tables when the number of rows is constant," *Proceedings of the 34th Annual ACM Symposium on Theory of Computing* (STOC 2002), pp. 240–249.

[11] M. Cryan, M. Dyer, L. A. Goldberg, M. Jerrum, and R. Martin, "Rapidly mixing Markov chains for sampling contingency tables with constant number of rows," *Proceedings of the 43rd Annual Symposium on Foundations of Computer Science* (FOCS 2002), pp. 711–720.

[12] P. Diaconis and B. Effron, "Testing for independence in a two-way table: new interpretations of the chi-square statistic (with discussion)," *The Annals of Statistics*, **13** (1985), pp. 845–913.

[13] P. Diaconis and L. Saloff-Coste, "Random walk on contingency tables with fixed row and column sums," *Technical report*, Department of Mathematics, Harvard University, 1995.

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

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

[16] M. Dyer, A. Frieze, and R. Kannan, "A Random Polynomial Time Algorithm for Approximating the Volume of Convex Bodies," *Journl of the ACM*, **38** (1991), pp. 1–17.

[17] M. Dyer and C. Greenhill, "Polynomial-time counting and sampling of two-rowed contingency tables," *Theoretical Computer Sciences,* **246** (2000), pp. 265–278.

[18] M. Dyer, R. Kannan, and J. Mount, "Sampling contingency tables," *Random Structures and Algorithms*, **10** (1997), pp. 487–506.

[19] R. A. Fisher, "The logic of inductive inference (with discussion)," *Journal of Royal Statistical Society*, **98** (1935), pp. 39–54.

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

[21] V. Guruswami, "Rapidly Mixing Markov Chains: A Comparison of Techniques (A Survey)," 2000. (available from http://www.cs.washington.edu/homes/venkat/pubs/papers.html)

[22] O. Häggström, *Finite Markov Chains and Algorithmic Application*, London Mathematical Society, Student Texts **52**, Cambridge University Press, 2002.

[23] D. Hernek, "Random generation of $2 \times n$ contingency tables," *Random Structures and Algorithms*, **13** (1998), pp. 71–79.

[24] M. Jerrum and A. Sinclair, "The Markov chain Monte Carlo method: an approach to approximate counting and integration," In *Approximation Algorithm for NP-hard Problems* (Dorit Hochbaum, ed.), PWS, 1996, pp. 482–520.

[25] R. Kannan, L. Lovász, and M. Simonovitz, "Random Walks and an $O^*(n^5)$ Volume Algorithm for Convex Bodies," *Random Structures and Algorithms*, **11** (1997), pp. 1–50.

[26] D. Karger, "A Randomized Fully Polynomial Time Approximation Scheme for the All-terminal Network Rliablity Problem," *SIAM Jounal on Computing*, **29** (1999), pp. 492–514.

[27] S. Kijima and T. Matsui, "Approximate counting scheme for $m \times n$ contingency tables ," *METR* 2003-01, Mathematical Engineering Technical Reports, University of Tokyo, 2003. (available from http://www.keisu.t.u-tokyo.ac.jp/Research/techrep.0.html)

[28] S. Kijima and T. Matsui, "Polynomial Time Perfect Sampling Algorithm for Two-rowed Contingency Tables," *METR* 2003-15, Mathematical Engineering Technical Reports, University of Tokyo, 2003. (available from http://www.keisu.t.u-tokyo.ac.jp/Research/techrep.0.html)

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

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

[31] T. Matsui and S. Kijima, "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)

[32] T. Matsui, M. Motoki and N. Kamatani, "Polynomial Time Approximate Sampler for Discretized Dirichlet Distribution," *METR* 2003-10, Mathematical Engineering Technical Reports, University of Tokyo, 2003. (available from http://www.keisu.t.u-tokyo.ac.jp/Research/techrep.0.html)

[33] T. Matsui, Y. Matsui, and Y. Ono, "Random generate of $2 \times \cdots \times 2 \times J$ contingency tables," *METR* 2003-03, Mathematical Engineering Technical Reports, University of Tokyo, 2003. (available from http://www.keisu.t.u-tokyo.ac.jp/Research/techrep.0.html)

[34] Mersenne Twister Home Page, http://www.math.keio.ac.jp/~matumoto/mt.html

[35] 元木光雄, 鎌谷直之, 「Dirichlet 分布に従う多項分布の母数の Makov chain を用いた approximate sampler」, 2001 年冬の LA シンポジウム.

[36] R. Motowani and P. Raghavan, *Randomised Algorithms*, Cambridge University Press, Cambridge, UK, 1995.

[37] T. Niu, Z. S. Qin, X. Xu and J. S. Liu, "Bayesian haplotype inference for multiple linked single-nucleotide polymorphisms," *American Journal of Human Genetics*, **70** (2002) 157–169.

[38] 大森裕浩, "マルコフ連鎖モンテカルロ法の最近の展開," 日本統計学会誌, **31** (2001), pp. 305–344.

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

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

[41] J. Propp and D. Wilson, "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), pp. 170–217.

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

[43] 坂田年男, 澤江隆一, "計算機統計学と数学の連携 (分割表の解析の最近の話題から) ," 日本計算機統計学会第 *16* 回大会論文集 (2002), pp. 92–101.

[44] A. Sinclair, "Improved bounds for mixing rates of Markov chains and multicommodity flow," *Combinatrics, Probability and Computing*, **1** (1992), pp. 351–370.

[45] A. Sinclair, *Algorithms for Random Generation & Counting: A Markov Chain Approach*, Birkhäuser, Boston, 1993.

[46] A. Sinclair and M. Jerrum, "Approximate counting, uniform generation and rapidly mixing Markov chains," *Information and Computation*, **82** (1989), pp. 93–133.

[47] L. G. Valiant, "The complexity of computing the permanent," *Theoretical Computer Science*, **8** (1979), pp. 189–201.

[48] V. V. Vazirani, *Approximation Algorithms*, Springer-Verlag Berlin Heidelberg, New York, 2003.

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

[50] Web Site for Perfectly Random Sampling with Markov Chains, http://research.microsoft.com/~dbwilson/exact/

付 録A

Read Onceアルゴリズム

この章では2000年にWilsonの提案したRead onceアルゴリズムについて述べる[49]。このアルゴリズムは特殊なCFTPアルゴリズムとみなすことができ、乱数の列を記憶することなくパーフェクトサンプリングが行える。以下、有限離散の状態空間Ωとupdate functionϕをもつエルゴード的なマルコフ連鎖$\mathcal{M}$に対するアルゴリズムを考える。アルゴリズム中では第2章で述べたCTTF (Coupling To The Future) を利用する。まず準備として、アルゴリズム中で呼び出す2つのプロセス、*-successfulと*-failingについて説明する。

プロセス A (*-successful)

Step 1. 2つの空列$\boldsymbol{\lambda}_1$, $\boldsymbol{\lambda}_2$を用意する。シミュレーションの終了時刻を$T := 1$に設定する。

Step 2. 2つの実数乱数$\lambda_1[T-1]$, $\lambda_2[T-1]$を生成し、それぞれ$\boldsymbol{\lambda}_1$, $\boldsymbol{\lambda}_2$の最後尾に加える。すなわち$\boldsymbol{\lambda}_1 = (\lambda_1[0], \ldots, \lambda_1[T-2], \lambda_1[T-1])$, $\boldsymbol{\lambda}_1 = (\lambda_2[0], \ldots, \lambda_2[T-2], \lambda_2[T-1])$とする。

Step 3. 数列$\boldsymbol{\lambda}_1$と$\boldsymbol{\lambda}_2$を用いて2本のCTTFを実行する。

Step 4. coalesceence checkを行う。

(a) $\boldsymbol{\lambda}_1$と$\boldsymbol{\lambda}_2$のどちらか片方でもcoalesceしていなければ$T := T + 1$としてStep 2に戻る。

(b) $\boldsymbol{\lambda}_1$と$\boldsymbol{\lambda}_2$ともにcoalesceした場合。先にcoalesceした乱数列の添え字$i \in \{1, 2\}$とする。同時の場合は$1/2$の確率でiを決める。数列$\boldsymbol{\lambda}_i$に対する状態すなわち$\Phi_0^T(x, \boldsymbol{\lambda}_i)$の値を返して終了する。

プロセス B (*-failing)

Input: 状態$x \in \Omega$を入力する。

Step 1. 2つの空列$\boldsymbol{\lambda}_1$, $\boldsymbol{\lambda}_2$を用意する。シミュレーションの終了時刻を$T := 1$に設定する。

Step 2. 2つの実数乱数$\lambda_1[T-1]$, $\lambda_2[T-1]$を生成し、それぞれ$\boldsymbol{\lambda}_1$, $\boldsymbol{\lambda}_2$の最後尾に加える。すなわち$\boldsymbol{\lambda}_1 = (\lambda_1[0], \ldots, \lambda_1[T-2], \lambda_1[T-1])$, $\boldsymbol{\lambda}_1 = (\lambda_2[0], \ldots, \lambda_2[T-2], \lambda_2[T-1])$とする。

Step 3. 数列$\boldsymbol{\lambda}_1$と$\boldsymbol{\lambda}_2$を用いて2本のCTTFを実行する。

Step 4. coalesceence checkを行う。

(a) $\boldsymbol{\lambda}_1$と$\boldsymbol{\lambda}_2$がともにcoalesceしていなければ$T := T + 1$としてStep 2に戻る。

(b) $\boldsymbol{\lambda}_1$と$\boldsymbol{\lambda}_2$のいずれかがcoalesceした場合。coalesceしていない乱数列の添え字を$j \in \{1, 2\}$とする。同時の場合は$1/2$の確率でjを決める。初期状態xと数列$\boldsymbol{\lambda}_j$に対する状態すなわち$\Phi_0^T(x, \boldsymbol{\lambda}_j)$の値を返して終了する。

以上の準備を用いてRead Onceアルゴリズムを以下に記述する。

アルゴリズム **4**

Step 1. プロセス A (*-successful) を実行し、変数 $x \in \Omega$ に出力された値を代入する。$1/2$ の確率で次の (a), (b) のいずれかを実行する。

(a) $x \in \Omega$ を出力しアルゴリズムを終了する。

(b) Step 2 に進む。

Step 2. 入力 $x \in \Omega$ を与えてプロセス B (*-failing) を実行し、変数 $x \in \Omega$ を出力された値に更新する。$1/2$ の確率で次の (a), (b) のいずれかを実行する。

(a) $x \in \Omega$ を出力しアルゴリズムを終了する。

(b) Step 2 にもどる。

定理 **A.1** アルゴリズム 4 が確率 1 で終了するならば、出力された値はマルコフ連鎖 $\mathcal{M}$ の定常分布に厳密に従う。 ■

アルゴリズム 4 によって、次の CFTP アルゴリズムが実現されることになる。

アルゴリズム **5**

Step 1. シミュレーションの開始時刻 $T = 0$ と空列 $\boldsymbol{\lambda}$ を用意する。反復回数を $i = 1$ にする。

Step 2. 1 本の CTTF を実行し、その coalescence time を $T_i > 0$ とする。

Step 3. シミュレーションの開始時刻を $T := T - T_i$ とする。一様実数乱数 $\lambda[T], \ldots, \lambda[T + T_i - 1]$ を生成して、$\boldsymbol{\lambda} = (\lambda[T], \ldots, \lambda[-1])$ とする。数列 $\boldsymbol{\lambda}$ を用いて CFTP を時刻 T から 0 まで実行する。(このとき、$1 - 1/2^i$ 以上の確率で CFTP アルゴリズムは coalesce していることに注意。)

Step 4. 確率 $1/2$ で、CFTP シミュレーションの時刻 0 での値を返し、アルゴリズムを終了する。残りの場合、$i := i + 1$ として Step 2 に戻る。

付 録B

計算機実験

本章では、提案したCFTPアルゴリズムにおけるcoalescence timeについて計算機実験を行い、本文中3.3節および4.3節で得られた計算時間の算定の評価を行う。第2章の2.2節で示したように、CFTPのcoalescence timeとCTTFのcoalescence timeは同一の分布に従う。一方、CFTPのcoalescence timeの厳密な算定にはcoalescence timeの2乗のオーダーの計算時間がかかるのに対して、CTTFの場合、線形オーダーの計算時間で計算可能である。このことは定理3.6および定理4.5に反しないことに注意が必要である。

以上のことを踏まえて、提案した単調マルコフ連鎖に対してCTTFを行い、coalescence timeについて計算機実験を行った。計算機実験の環境は以下のとおりである。

CPU	Pentium 4 1.7GHz
メモリ	512MB
OS	Windows XP Professional
開発環境	Visual C++ 6.0

以下、実験では各インスタンスに対し10,000回の試行を行い、coalescence timeの平均値をとった。また、乱数の生成にはMersenne Twister ([34]) を使用した。

B.1 2行分割表のCoalescence Time

この節では2行分割表のパーフェクトサンプリング法に関する実験について述べる。まず、条件3.1の必要性を確認するために、多項式性を損なうと考えられるケースについて実験を行った。次の実験では列和の順番のみが異なる2つの場合について実験を行った。1つは行和$\boldsymbol{r} = (500, 500)$と列和$\boldsymbol{s} = (500, 1, 1, 498)$の分割表の集合で、もう1つは条件3.1を満たすように並び替えた行和$\boldsymbol{r} = (500, 500)$と列和$\boldsymbol{s} = (500, 498, 1, 1)$の分割表の集合である。明らかにこの2つの分割表集合は同等である。次の表B.1はこの問題に対して実験を行ったものである。この実験の結果は条件3.1が計算時間の多項式性において本質的な条件であることを示唆する。

表B.1. 特殊データに対するcoalescence timeの平均値。データの単位は回。

列和ベクトル$\boldsymbol{s}$ (条件3.1)	平均推移回数
$(500, 1, 1, 498)$ (満たさない)	2105440.8000
$(500, 498, 1, 1)$ (満たす)	21.1924

次に列数 n (n:偶数) で、行和 $\boldsymbol{r} = (n/2, n/2)$ と列和 $\boldsymbol{s} = (1, 1, \ldots, 1)$ を持つ分割表の集合について実験を行った。この場合、分割表の合計値 N は $N = n$ となるため、期待値の理論上限は $\mathrm{O}(n^3\ lnn)$ である。この条件で coalescence time の理論的期待値について計算機実験を行った。表 B.2 はその結果である。また図 B.1 は表 B.2 の結果をグラフにしたものである。

表 B.2. 列数 n と coalescence time の関係。データは coalescence time の平均値で、単位は回。

列数 n	平均推移回数
$n = 10$	455.2997
$n = 20$	4779.2388
$n = 30$	18530.1556
$n = 40$	47493.0267
$n = 50$	98986.6891
$n = 60$	179504.2158
$n = 70$	294804.5399
$n = 80$	454908.1312
$n = 90$	665046.3705
$n = 100$	929777.5032

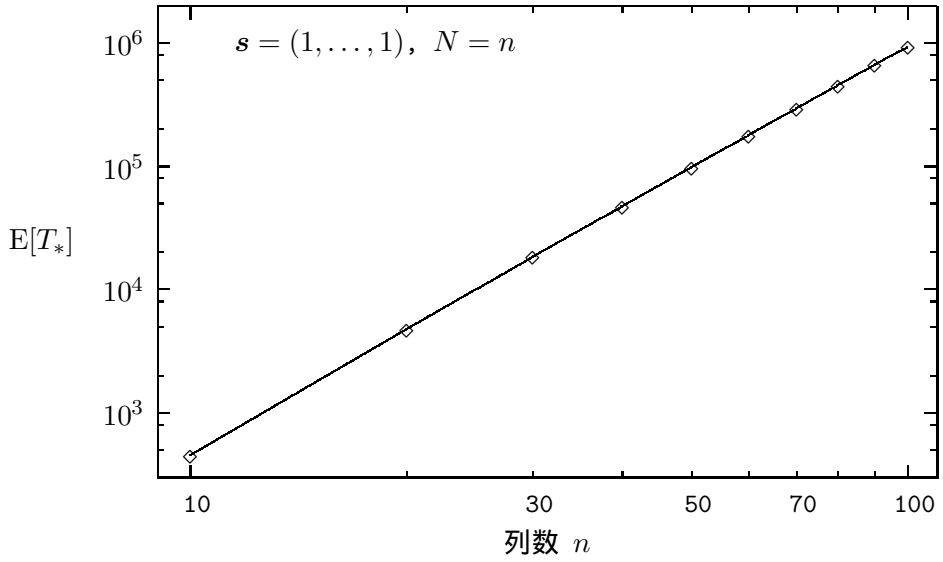


図 B.1. 列数 n と平均 coalescence time $\mathrm{E}[T_*]$ のグラフ。

次に表の合計値 N と coalescence time に関する実験を行った。この実験で用いた問題は表の合計値が N、行和が $(N/2, N/2)$、列和が $(N/10, N/10, \ldots, N/10)$ である。表 B.3 はその結果である。また図 B.2 は表 B.3 の結果をグラフにしたものである。

表 B.3. 表の合計値 N と coalescence time の関係。データは coalescence time の平均値で、単位は回。

合計値 N	平均推移回数
$N = 10$	455.2997
$N = 30$	635.7263
$N = 100$	846.1706
$N = 300$	1036.0252
$N = 10^3$	1240.2445
$N = 10^4$	1637.4041
$N = 10^5$	2033.1881
$N = 10^6$	2428.0224
$N = 10^7$	2825.2828
$N = 10^8$	3218.1437
$N = 10^9$	3610.3090

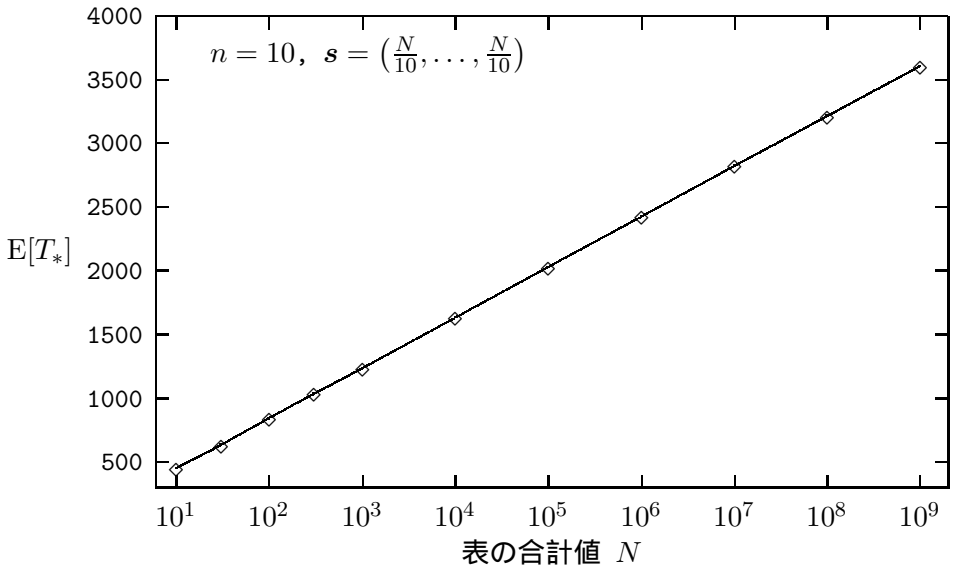


図 B.2. 表の合計値 N と平均 coalescence time $\mathrm{E}[T_*]$ のグラフ。

最後に、さまざまな分割表の列数 n および合計値 N に対して、行和と列和をランダムに生成したインスタンスを用意して実験を行った。表 B.4 はこの結果をまとめたものである。表の値は、各インスタンスに対して 10,000 回の試行を行い、coalescence time の平均値を取ったものである。また、括弧内は Read Once アルゴリズム (付録 A 参照) を実行し、10,000 サンプルを得るのに要した時間で、単位は秒である。

表 B.4. ランダムデータに対する coalescence time の平均値。データは coalescence time の平均値で、単位は回。(括弧内は実際に 10,000 サンプルを得るのに必要な時間で、単位は秒。)

		合計値 N		
		$N = 10^3$	$N = 10^4$	$N = 10^5$
列数 n	$n = 10$	1157.3895 (48.593)	1599.8796 (67.187)	1767.7371 (73.734)
	$n = 15$	4265.5681 (176.859)	5016.5844 (209.718)	5908.3854 (244.156)
	$n = 20$	9642.4697 (395.515)	13473.2172 (565.171)	14444.6369 (598.109)
	$n = 25$	16007.2848 (675.921)	26803.3977 (1115.031)	21964.875 (911.328)
	$n = 30$	27739.6633 (1166.672)	44529.6742 (1858.515)	46197.3365 (1938.984)
	$n = 40$	77620.9302 (3282.718)	85036.9358 (3614.562)	106695.5401 (4425.656)
	$n = 50$	161168.3618 (6764.234)	141916.2711 (5988.828)	233225.26320 (9865.125)

B.2　離散化 Dirichlet 分布の Coalescence Time

この節では離散化 Dirichlet 分布に従うパーフェクトサンプリング法に関する実験について述べる。まず、離散化幅と coalescence time の関係について調べた。Dirichlet 次元 $n = 10$、Dirichlet パラメータを $(0.5, \ldots, 0.5)$ に設定し実験を行った。表 B.5 はその結果である。また図 B.3 は表 B.5 の結果をグラフにしたものである。

表 B.5. 離散化サイズ Δ と coalescence time の関係。データは coalescence time の平均値で、単位は回。

離散化幅Δ	平均推移回数
$\Delta = 100$	972.3839
$\Delta = 200$	1089.5130
$\Delta = 300$	1154.8049
$\Delta = 400$	1199.7975
$\Delta = 500$	1234.1825
$\Delta = 600$	1264.9268
$\Delta = 700$	1288.4943
$\Delta = 800$	1308.7422
$\Delta = 900$	1331.0509
$\Delta = 1000$	1346.4733

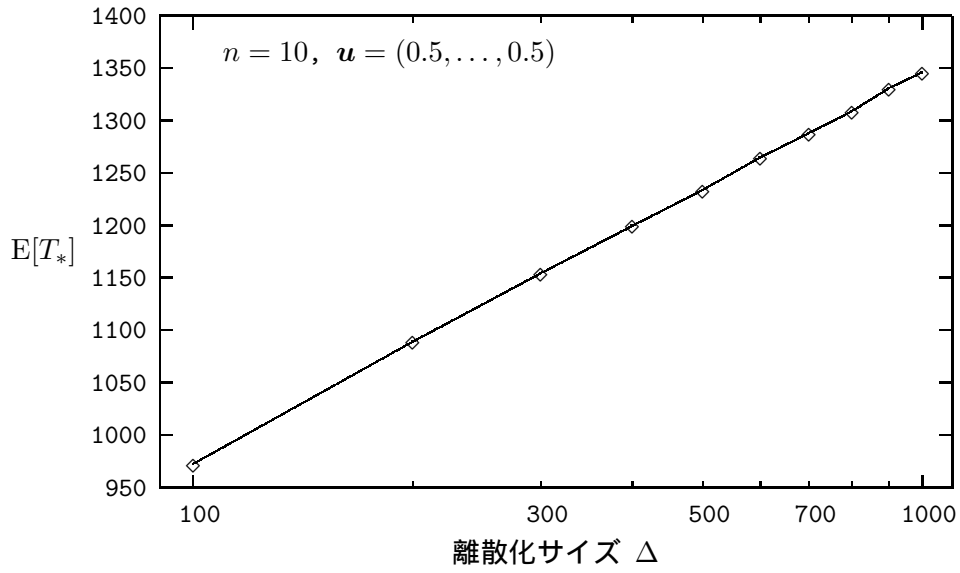


図 B.3. 離散化幅 Δ と平均 coalescence time $\mathrm{E}[T_*]$ のグラフ。

最後に Dirichlet 次元 n と coalescence time の関係について調べた。Dirichlet パラメータを $(0.5, \ldots, 0.5)$、離散化幅を $\Delta = 400$ に設定し実験を行った。表 B.6 はその結果である。また図 B.4 は表 B.6 の結果をグラフにしたものである。

表 B.6. Dirichlet 次元 n と coalescence time の関係。データは coalescence time の平均値で、単位は回。

列数 n	平均推移回数
$n = 10$	1199.7975
$n = 15$	4334.8648
$n = 20$	10630.1277
$n = 30$	36776.8255
$n = 40$	88167.1974
$n = 50$	173121.9811
$n = 60$	299131.3983
$n = 70$	475907.4327
$n = 80$	707257.5348
$n = 90$	1002441.1870
$n = 100$	1382016.3669

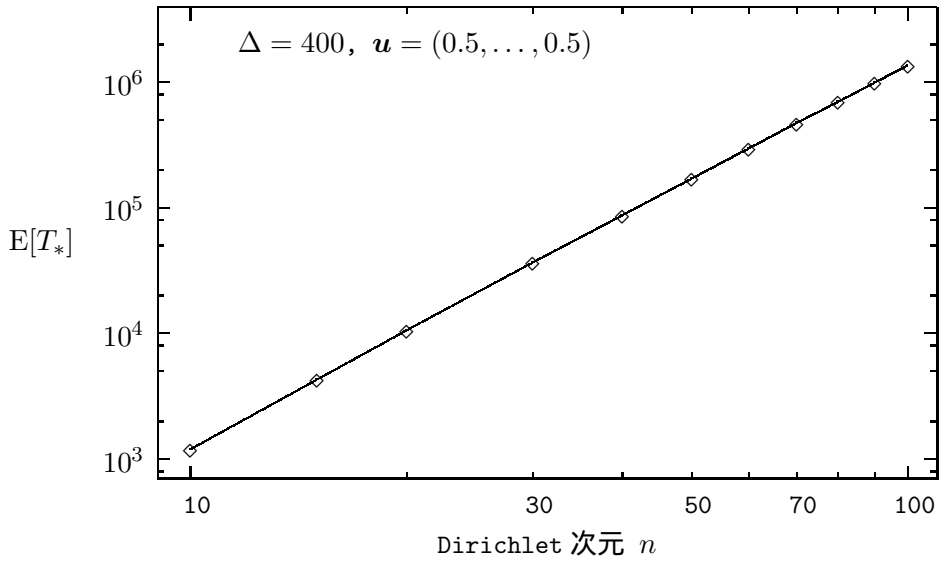


図 B.4. Dirichlet 次元 n と平均 coalescence time $\mathrm{E}[T_*]$ のグラフ。