MCMC の基礎と画像解析への応用

坂田　年男　　九州大学　芸術工学研究院

1　はじめに

近年、さまざまな分野で応用されているマルコフ連鎖モンテカルロ法（ＭＣＭＣ）の基礎と画像解析への応用について解説する。ＭＣＭＣの基礎については、すでにさまざまな解説がある。例えば、統計学のフロンティアシリーズに、計算統計ＩＩ（マルコフ連鎖モンテカルロ法とその周辺（伊庭幸人他 (岩波)) なる comprehensive な解説書 [6] が出たばかりであり、また、大森裕浩先生 [7] の日本統計統学会誌における「マルコフ連鎖モンテカルロ法の最近の発展」というきわめて充実した解説もある中ではあるが、本稿では画像関係でＭＣＭＣを使って精力的に仕事をされているバークレーの S.C.Zhu 教授の仕事が容易に理解できるような材料を提供することに狙いを定め、Trans-Dimensional MCMC [3, 4]、Graph Partitioning [1], Data Driven MCMC [10] 等の解説に焦点を当てて解説する。

2　マルコフ連鎖モンテカルロ法の基礎

時間と共に有限な状態空間 Ω を彷徨するランダムな要素 $X_t, t = 0, 1, 2,,,$ がマルコフ連鎖であるとは $x = x_{t-1}$ として、

$$P(x, y) = P\{X_t \in y | X_{t-1} = x_{t-1}, \ldots, X_0 = x_0\} = P\{X_t = y | X_{t-1} = x_{t-1}\} \tag{1}$$

が成り立つ事をいう。ここで、$P(x, y)$ をマルコフ連鎖 $\{X_t\}$ の推移確率 (行列) という。上の式は X_t の確率的振る舞いが 1 時刻前の状態 X_{t-1} のみにしか依存しない事を述べたものである。また、同時に、推移確率が時刻 t に依存しないことも述べており、時間的に homogeneous なマルコフ連鎖の定義となっていることに注意されたい。このとき、n ステップ後の推移確率は

$P_n(x, y) = P(X_{t+n} = y | X_t) = [P]^n(x, y)]$ で与えられる。右辺は行列の n ベキである。ここで、状態空間を有限と仮定するのは理論および表記が簡単になるためである。Ω の上の確率分布 μ に対して、Ω 上の確率 μP を

$$\mu P(y) = \sum_x \mu(x) P(x, y) \tag{2}$$

と定める。$\mu P^t = (\mu P^{t-1})P, t = 1, 2,,,$ によって、初期分布 μ から出発した t 時刻における $\{X_t\}$ の確率分布が求まる。Ω 上の確率 π は $\pi P = \pi$ を満たす時、マルコフ連鎖 $\{X_t\}$ の定常分布といわれる。ＭＣＭＣ法とは Ω 上に確率分布 π があり、π から直接サンプルを取り出すことが困難なときに, 適当な初期分布 μ から出発するマルコフ連鎖で μP^t が極限 $t \to \infty$ において、定常分布 π に収束

するマルコフ連鎖 $\{X_t\}$ を構成し、適当な時刻で定常分布 π へ収束したとみなし、それ以後の X_t を定常分布 π からのサンプルとして利用する方法である。ここで理論的に問題となるのは、

(1) 定常分布はどのような条件の下に存在するか?

(2) 特定のマルコフ連鎖の t 時刻での分布が極限で定常分布に収束することはどのような条件があれば保障されるか?

(3) 定常分布は初期分布に無関係に定まるか？

(4) 収束の早さはどの程度か?

(5) サンプルを使って期待値計算などが出来るのか?

などである。また、実際問題としては

(6) どのようにして、目的分布 π を定常分布にもつマルコフ連鎖を構成するか？

(7) 収束したことをどのように判定するか?

が問題となる。(1),(2),(3),(4) に対しては、マルコフ連鎖に関する確率論の以下の定理によって、問題は解決する。若干用語を準備すると、マルコフ連鎖が既約であるとは任意の Ω の 2 点が互いに到達可能であることをいう。既約なマルコフ連鎖において、各状態 x に対して、x から出発して x に戻ることができるステップ数の最大公約数を $d(x)$ で表し、状態 x の周期という。すべての x に対して、$d(x) = 1$ のとき、マルコフ連鎖は非周期的であるという。

定理 1. 有限集合上のマルコフ連鎖が既約で非周期的であれば、ただひとつの定常分布 π を持ち、すべての状態 x, y に対して、$\lim_{t\to\infty} P_t(x,y) = \pi(y)$ がなり立つ。さらに、すべての状態 x, y とすべての時刻 t に対して、

$$P_t(x,y) - \pi(y)| \leq cr^t \tag{3}$$

が成り立つ定数 $c > 0$ と $0 < r < 1$ が存在する。

定理 1 より収束は指数的であることまで保障されていることに注意。(5) の問題は「マルコフ連鎖に対する大数の法則が成り立つか」という問題である。それは次の定理によって解決するが、そのために用語を準備する。Ω 上の二つの確率分布 μ, ν に対して、全変動距離 $\|\cdot\|$ を

$$\|\mu - \nu\| = max_{x,y\in\Omega}|\mu(x) - \nu(x)| \tag{4}$$

と定め、推移行列 $P(x,y)$ に対して、contraction 係数 $c(P)$ を

$$c(P) = \frac{1}{2} max_{x,y\in\Omega}\|P(x,) - P(y,)\| \tag{5}$$

と定義する。このとき、次の定理が成り立つ。

定理 2 マルコフ連鎖が定常分布 π をを持つとする。さらに、$c(P) < 1$ とする。このとき、任意の Ω 上の関数 $f(x)$ に対して、任意の初期分布から出発する X_t に対して、

$$\lim_{n\to\infty} \frac{1}{T} \sum_{t=1}^{T} f(X_t) = E_\pi\{X\} \tag{6}$$

が $L^2(P_\mu)$ の意味で成立する。ここで、$L^2(P_\nu)$ の意味での収束とは、$T \to \infty$ のとき、

$$E_\mu|\frac{1}{T}\sum_{i=1}^{T} f(X_i - E_\pi\{X\}|^2 \to 0 \tag{7}$$

が成り立つことをいう。
さらに、任意の ϵ に対して、定数 $C = C(\epsilon)$ が存在し

$$P\{|\frac{1}{n}\sum_{t=1}^{T} f(X_t) - E_\pi(f)| > \epsilon)\} \leq \frac{C}{(1-c(P))T\epsilon^2} \tag{8}$$

が成り立つ。
定常分布が目標とする分布 π であるマルコフ連鎖を構成する一般的方法としてメトロポリス-ヘステイング法がある。この枠内にギッブスサンプリング、シミュレーティッツドアニーリング法などがある。メトロポリス-ヘステイング法とは、現在の位置を $X_t = x$ とするときに、適当なマルコフ核 $Q(x,)$ から次の位置 y をサンプリングし、確率

$$r = \max\{1, \frac{\pi(y)Q(y,x)}{\pi(x)Q(x,y)}\} \tag{9}$$

で $X_{t+1} = y$ に推移し, 確率 $1-r$ で $X_{t+1} = x$ に留まるマルコフ連鎖を発生させる手法である。このマルコフ連鎖が目標とする分布 π を定常分布として持つことは、このマルコフ連鎖の推移核を $K(x,y)$ とするとき、詳細つりあい条件

$$\pi(y)K(x,y) = \pi(y)K(y,x) \tag{10}$$

を満たすことから容易に示される。ギッブスサンプリング法は密度関数が $p(x_1,...,x_p)$ の p 変量分布からのサンプリングが困難なときに、1 次元条件付き分布からのサンプリングを繰り返し行う事によって、p 変量分布からのサンプリングを行うもので、以下のアルゴリズムとなる。
(1) 初期値 $x^{(0)} = (x_1^{(0)},...,x_p^{(0)})$ を設定する。
(2) $X_1 = x_1^{(1)}$ を条件付分布 $p(x_1|X_2 = x_2^{(0)},...,X_p = x_p^{(0)})$ から発生させる。
(3) $X_2 = x_2^{(1)}$ を条件付分布 $p(x_2|X_1 = x_1^{(1)}, X_3 = x_3^{(0)},...,X_p = x_p^{(0)})$ から発生させる。
(4) $X_3 = x_3^{(1)}$ を条件付分布 $p(x_3|X_1 = x_1^{(1)}, X_2 = x_2^{(1)}, X_4 = x_4^{(0)}...,X_p = x_p^{(0)})$ から発生させる.
...
(5) $X_p = x_p^{(1)}$ を条件付分布 $p(x_p|X_1 = x_1^{(1)}, X_2 = x_2^{(1)},...,X_{p-1} = x_{p-1}^{(1)})$ から発生させる。
(6) (1) において、初期値を $x^{(1)} = (x_1^{(1)},...,x_p^{(1)})$ とおいて (2) から (6) を繰り返す。
(7) (6) を初期値を生成した x_n と置き換えて繰り返す。
このとき、マルコフ連鎖 $x^{(0)} \to x^{(1)} \to x^{(2)} \to \cdots$ は p 次元空間の既約かつ非

周期的な密度 $p(x_1,...,x_p)$ に関して詳細つりあい条件を満たすマルコフ連鎖となるので、その定常分布が目標である分布 $p(x_1,...,x_p)$ となるという仕組みである。一方、シミュレーティッドアニーリングの方は最大化問題に対するランダム探索手法の一つである。$f(x_1,...,x_p)$ を最大化したい関数とするとき、ギッブス分布

$$p(x|\beta) = \frac{exp(\beta f(x))}{\sum_x exp(\beta f(x))} \tag{11}$$

に対して、$\beta = \beta_1$ とおき、初期値 $x^{(0)}$ から出発するマルコフ連鎖モンテカルロ法を使って、ギッブス分布 $p(x|\beta_1)$ からのサンプリング $x_1^{(1)},...,x_{n_1}^{(1)}$ を行い、次に $\beta = \beta_2$ とおき、最後の $x_{n_1}^{(1)}$ を初期値として、ギッブス分布 $p(x|\beta_2)$ からのサンプリングを行う。これを繰り返していく手法である。$\beta \to \infty$ のとき、$x^{(0)} \to x^{(1)} \to x^{(2)} \to \cdots$ は $f(x)$ の最大値を与える点に収束する。

3 画像解析のためのMCMC

画像解析における MCMC の世界的リーダーの一人はバークレーの S.C.Zhu 教授ではなかろうか？　彼の論文のキーワードを年代順に追っただけでも、Forms [12], Frame [13],Minimax Entropy Principle [14], Generalized Swendsen-Wang Cuts [1], Data Driven Markov Chain [10], など、その活躍ぶりは目を見張るものがあるが、彼の仕事のモチーフは一貫してベイズ理論による画像の確率・統計的解釈にあると云えよう。これは、変分原理による画像解釈を追求するミネソタ大学の Jackie Shen 教授 [5] と好対照である。この節では特に、S.C.Zhu 教授の二つの論文 Graph Partition by Swendsen-Wang Cuts [1] と Image Segmentation by Data Driven Markov Chain Monte Carlo [10] について簡単に解説する。

3.1 Graph Partition by Swendsen-Wang Cut

この論文はイメージ解析の主要な目的の一つである自然画像における物体の分離、抽出、をグラフ分割問題として捉え、そこに、Trans-Dimensional MCMC を適用したものである。具体的に説明しよう。$G_0 = \langle V, E_0 \rangle$ を隣接グラフとする。そこで、頂点集合 $V = \{v_1,...,v_N\}$ はピクセルやエッヂレットやプリミティブなどのようなものを考えており、$E_0 = \{e = \langle s,t \rangle\}$ を辺集合とする。目的は G_0 を未知の個数 n の完全部分グラフ $G_k = \langle V_k, E_k \rangle, k = 1,2,,,n$ に分解する事である。ここで、辺集合 E_k は G_0 の辺をそのまま踏襲したものである。また、頂点集合 V_k は V の交わりの無い分割をなすものとし、そのような n 分割を一般に π_n で表す。Ω_{π_n} でそのようなすべての n 分割の全体を表わすとする。目標は各頂点集合 V_k を画像として一貫しているように分割することである。すべての可能な分割の仕方の空間は $\Omega_\pi = \cup_{n=1}^{|V|}\Omega_{\pi_n}$ である。頂点集合の対 V_i, V_j に対して、$C(V_i,V_j) = \{e = \langle s,t \rangle | e \in E_0, s \in V_i, t \in V_j\}, i \neq j$ を cut という。アル

ゴリズムを説明するための準備をする。一般にグラフ $G = \langle V, E_0 \rangle$ において、各辺 e を確率 $q_e, 1-q_e$ で独立に $\{on, off\}$ に切り替えることで、$G = \langle V, E \rangle$ が生成される。そこで、$E \subset E_0$ は偶然 {on} となった E の辺の集合を表す。このとき、

$$q(E) = \prod_{e \in E} q_e \prod_{e \in E_0 - E} (1 - q_e)$$

となることが分かる。このとき、$G = \langle V, E \rangle$ の連結成分 $G_k = \langle V_k, E_k \rangle$、$k = 1, \ldots, n$ への分解を考え、$CP = \{V_1, \ldots, V_n\}$ とおく。Swendsen-Wang Cuts アルゴリズムは以下のとおりである。

$G_0 = \langle V, E_0 \rangle$ と $\{q_e | e \in E_0\}$ と事後確率 $p(W|I)$ を入力し $W \sim p(W|I)$ をサンプリングし出力するのが目的。

1. グラフ分割 $\pi : \cup_{i=1}^{n} G_i$ を初期分割とする。
2. 現在の分割を A とする。
3. 各部分グラフ $G_i = \langle V_i, E_i \rangle$ において、すべての辺 $e \in E_i$ を確率 $q_e, 1-q_e$ で $\{on, off\}$ する。
4. V_i を n_i 個の連結成分に分解し、連結部分グラフ $G_{ij} = \langle V_{ij}, E_{ij} \rangle, j = 1, \ldots, n_i$ を生成する。$CP = \{V_{ij},\ \ i = 1, \ldots, n, j = 1, \ldots, n_i\}$ とおく。
5. $V_0 \in CP$ を確率 $q(V_0|CP)$ で取り出す（通常は確率 $1/|CP|$ で選ぶ)。
6. $\{1, 2, \ldots, n+1\}$ 上の確率分布 $q(\ell|V_0, A, G_0)$ を

$$q(\ell|V_0, A, G_0) = \left\{ \begin{array}{ll} a & G_\ell \text{ が } V_0 \text{ に隣接している場合} \\ b & \ell = n+1 \text{ の場合} \\ c & \text{その他の場合} \end{array} \right)$$

で定め (適当な a,b,c)、V_0 を部分グラフ G_ℓ に割り付けることを提案する。$\ell \leq n$ のときは V_0 を G_ℓ に吸収させ、その時のグラフ分割状態を B で表す。$\ell = n+1$ の場合は V_0 を新たな部分グラフとして、その時に出来るグラフの分割状態を C と書く。
7. 現在のグラフ分割状態 A から B または C への推移確率 $\alpha(A \to B), \alpha(A \to C)$ を適切に定め推移させるか否かを確率的に定める.
8. 2 から 7 を繰り返す。

この時、全体としてのマルコフ連鎖は ergodic で詳細つりあい条件を満たす事が示されるのである。

3.2 Image Segmentation by Data Driven Markov Chain Monte Carlo

この論文はイメージ分割問題をグラフでなくユークリッド (lattice) 空間で直接扱っている。パラメータ空間は階層的で、その上の Trans-Dimensional Markov Chain を構成している。連鎖の収束のスピードアップのために Data Driven Markov Chain を適用しているのが大きな特徴となっている。少し具体

的に解説しよう。$\Lambda = \{(i,j)|1 \le i \le L, 1 \le j \le H\}$ をイメージ lattice とし、I_Λ を Λ 上のイメージとする。$I_v \in \{1,2,...,G\}, v \in \Lambda$ はグレーイメージならピクセル強度を表すと考える。 イメージ分割問題は Λ を未知個数 K に分割する事である。

$$\Lambda = \cup_{i=1}^{K} R_i,\ R_i \cap R_j = \phi,\ i \neq j.$$

領域 R の境界を ∂R で表す。領域 R のイメージ I_R は確率モデル $p(I_R, \Theta)$ の実現値とモデル化する。この確率モデルは添え字 ℓ でクラス分けされている。したがって、イメージを生み出すと考える確率モデルのパラメータは、

$$W = (K, (R_1, \ell_1, \Theta), ..., (R_K, \ell_K, \Theta_K))$$

であり、事前分布 $p(W)$ はかなり複雑で、細かい説明は省くが、

$$P(W) = exp(-\lambda_0 K - \sum_{i=1}^{K}[\mu \int_{\partial R_i} ds + \gamma|R_i|^c + \mu|\Theta_i|])$$

ととる。尤度の方は $p(I|W) = \prod_{i=1}^{K} p(I_{R_i}; \Theta_i, \ell_i)$ となり、$\ell_i = 1,2,3,4$ で異なる領域画像生成モデルを表し,Θ_i の動く空間は $\omega_\theta = \omega_{g_1} \cup \omega_{g_2} \cup \omega_{g_3} \cup \omega_{g_4}$ となる。ここに、ω_{g_1} は各ピクセル強度 I_v が独立な正規分布に従うというモデルに含まれるパラメータの空間、ω_{g_2} はノンパラメトリックな強度ヒストグラムモデルに現れるパラメーター空間、ω_{g_3} はピクセル間の相互作用を捉える texture モデル:Frame に含まれるパラメータの空間、ω_{g_4} は 2D Bezier-spline モデルに含まれるパラメータ空間を表す。これらを合わせたパラメータ空間は

$$\Omega = \cup_{k=1}^{|\Lambda|} \Omega_k = \cup_{k=1}^{|\Lambda|}[\omega_{\pi_k} \times \omega_\theta \times \cdots \times \omega_\theta]$$

となる。ここで、ω_{π_k} は Λ の $k-$ 分割の全体である。この設定のもとに、I が観測されたとき、Ω 上の事後分布 $p(W|I)$ からのサンプリングを MCMC 法で実現するのが目的である。マルコフ連鎖は ergodic であること、すなわち、任意の初期条件 $W_0 \in \Omega$ から出発して有限時間内に任意の状態を訪れるようにしたい。このために、jump-diffusion　マルコフ連鎖を取り入れる。Diffusion によって同じ次元のパラメータ空間を move させ、jump によって、異なる次元のパラメータ空間へ推移させるのである。このとき、事後分布 $p(W|I)$ に関して詳細つりあい条件を持つように設定する事が出来、$p(W|I)$ を定常分布とするマルコフ連鎖となる。

3.3 Trans-Dimensional MCMC

Zhu による画像解析への応用では、次元の異なるパラメーター空間をさまようマルコフ連鎖が使われている。これは Green [3, 4] によって、統計学の世界へ導入された手法である。これによって、ベイズモデル選択の MCMC 解析が可能と

なった。$k \in \mathcal{K}$ をモデルを指定するパラメータ、$\theta_k \in \mathcal{R}^{n_k}$ をモデル k の下での分布のパラメータ、Y を観測ベクトルとする。$p(k), p(\theta_k|k)$ を prior とし、$p(Y|k, \theta_k)$ を尤度関数とする。ここでの目的は、事後分布 $p(k, \theta_k|Y)$ からの MCMC サンプリングを行うことである。このときの状態空間は $\Omega = \cup_{k\in\mathcal{K}}(\{k\} \times \mathcal{R}^{n_k})$ である。各状態は $x = (k, \theta_k)$ である。$\Omega \subset \mathcal{R}^d$ における分布 π を定常分布をメトロポリス-ヘスティングス法で極実現したい。そのためには、マルコフカーネル $P(x, dx)$ が任意のボレル集合 A, B に対して次の詳細つりあい条件を満たすように定めればよい。

$$\int_{(x,x')\in A\times B} \pi(dx)P(x, dx') = \int_{(x',x)\in A\times B} \pi(dx')P(x', dx) \tag{12}$$

$P(x, dx)$ を提案分布 $q(x, dx')$ と採択率 $\alpha(x, x')$ で構成するとき、詳細つりあい条件は

$$\int_{(x,x')\in A\times B} \pi(dx)q(x, dx')\alpha(x, x') = \int_{(x',x)\in A\times B} \pi(dx')q(x', x)\alpha(x', x) \tag{13}$$

となる。このとき、$\pi(dx)q(x, dx')$ は $\Omega \times \Omega$ 上のある対称な測度 μ に dominate され (Green [3],Tierney [9])、Radon-Nikodim 微分 $f(x, x')$ を持ち、上の条件は

$$\int_{(x,x')\in A\times B} \alpha(x, x')f(x, x')\mu(dx, dx') = \int_{(x',x)\in A\times B} \alpha(x', x)f(x', x)\mu(dx', dx) \tag{14}$$

となる。これが成り立つには、

$$\alpha(x, x') = min\{1, \frac{f(x', x)}{f(x, x')}\}$$

とすれば十分であることが分かる。

4 あとがき

1 章ではシナジ [8],Winkler [11]、2 章では Zhu [15] 等をお手本にした。AdaBoost と DDMCMC を使った画像分割と検出と認識を統合的に扱う話 Zhu 等 [16] についても、出来れば講演では触れたい。

参考文献

[1] Barbu A. and S.C. Zhu S.C., Graph Partition by Swendsen-Wang Cut, Proceedings of 9-th IEEE ICCV2003)2nd.Volume, 2003.

[2] Brooks S.P., Giudici P. and Ryden T.,Efficient construction of reversible jump MCMC proposal distributions,Journal of the Royal Statistical Society, B, 65, pp.47-48, 2003.

[3] Green P.J., Reversible jump Markov chain Monte Carlo computation and Bayesian model determination,Biometrika, 82, pp.711-32, 1995.

[4] Green P.J., Trans-dimensional Markov chain Monte Carlo, Highly Structured Stochatic Systems, pp179-198, 2003, Oxford Univ. Press.

[5] Chan T.F. and Shen J.J.,Image Processing and Analysis-Variational,PDE,wavelet, and stochastic methods,SIAM Publisher, Philadelphia, 2005.

[6] 伊庭幸人、種村正美、大森裕浩、和合肇、佐藤整尚、高橋明彦, 計算統計 II（マルコフ連鎖モンテカルロ法とその周辺), 統計科学のフロンティア 12, 岩波出版, 2005.

[7] 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,pp-305-344,2001.

[8] シナジ R.B. , マルコフ連鎖から格子確率モデルへ, Springer 東京 (今野紀雄、林俊一訳), 2001.

[9] Tierny L., A note on Metropolis-Hastings kernels for general state spaces, Ann. of Applied Prob.,8,pp.1-9,1998.

[10] Tu Z.W. and Zhu S.C., Image Segmentation by Data Driven Markov Chain Monte Carlo,IEEE Trans on Pattern Analysis and Machine Intelligence, 24, no.5, pp.657-673, 2002.

[11] Winkler G., Image Analysis, Random Fields and Markov Chain Monte Calro Methods,Springer,2003(2nd ed.).

[12] Zhu S.C. and Yuille A.L., FORMS: A Flexible Object Recognition and Modeling System, International Journal of Computer Vision, 20, no.3, pp.187-212, 1996.

[13] Zhu S.C., Wu Y.N. and D.B. Mumford D.B.,FRAME: Filters, Random field And Maximum Entropy:Towards a Unified Theory for Texture Modeling, International Journal of Computer Vision, 27(2), pp.1-20, 1998.

[14] Zhu S.C., Wu Y.N. and D.B. Mumford D.B.,Minimax Entropy Principle and Its Applications to Texture Modeling,Neural Computation, 9, no.8, pp.1627-1660, 1997.

[15] Zhu S.C., Dellaert F. and Tu Z., Markov Chain Monte Carlo for Computer Vision, Tutorial at ICCV 2005.

[16] Tu.Z., Chen X., Yuille, A.L and Zhu S.C., Image Parsing:Unifing Segmentation, Detection, and Recognition, vol.2,ICCV2003.