数理解析研究所講究録
第 1768 巻 2011 年 73-84

マルコフ連鎖モンテカルロ法のエルゴード性の解析

東京大学大学院数理科学研究科 鎌谷研吾 (Kengo KAMATANI)[*1*2]
Graduate School of Mathematical Sciences
University of Tokyo

概要

本論文では，マルコフ連鎖モンテカルロ (MCMC) 法の収束性の解析のレビューを行う．MCMC 法の解析には，実用的にはマルコフ連鎖のエルゴード性の理論を用いることが一般的である．一方，独立型メトロポリス-ヘイスティングス法に限れば．より古典的な議論によって正確な漸近分散などの計算が出来ることを示す．

1 モンテカルロ法，マルコフ連鎖モンテカルロ法

マルコフ連鎖モンテカルロ (MCMC) 法は数値積分手法の中でも乱数を用いた手法の一つである．積分を数値計算することは本来は決定論的であるが，モンテカルロ法はあえて確率空間を導入し，非決定論的に近似計算を行う．MCMC 法はその中でもマルコフ連鎖を用いることが特徴である．応用分野は科学のさまざまな分野に渡る．

当面は MCMC 法のことは後においておき，モンテカルロ法について説明する．可測空間 $(E,\mathcal{E})$ 上に確率分布 P と可測関数 $f:E\to\mathbf{R}^d$ があって，

$$P(f):=\int_{x\in E}f(x)P(dx)$$

を計算する必要があるが $P(f)$ の解析的な計算が困難とする．このとき，数列 $(x_i;i=0,\ldots,n-1)$ を用いて

$$P_n(f):=n^{-1}\sum_{i=0}^{n-1}f(x_i)$$

により $P(f)$ を近似することを考える．$P_n(f)$ が $P(f)$ に近くなるように，どのように数列 $(x_i;i\geq 0)$ をとるかが問題である．最適な数列 $(x_i;i\geq 0)$ を選択する試みは low-discrepancy 列の概念へ導かれるだろう．モンテカルロ法ではこの問題に直接取り組まず，幾分違った視点を取る．数列 $(x_i;i\geq 0)$ を，ある特定の分布 μ の一つの実現値であるとみなし，$(x_i;i\geq 0)$ の取り方の問題を，$(x_i;i\geq 0)$ の分布 μ の取り方に捉え直す．そのため，$P_n(f)$ は確率変数であるから，μ の実現値 $(x_i;i\geq 0)$ に従い $P(f)$ に近かったり遠かったり，実現値によって異なるものと解釈される．しかし，このような $P_n(f)$ のふるまいは，デタラメではなく，分布によって完全に規定される．高い確率で $|P_n(f)-P(f)|$ の値が小さいのであれば，$P_n(f)$ を $P(f)$ の近似として使ってもよいと考える．そのため解析対象は数列ではなく，分布の方であり，確率論の道具を使って近似誤差などの様々な手法が使える．

例えば次のような積分 $\int_0^1(x+x^2)dx$ を考えよう．無論我々は答えを知っているわけだが，モンテカルロ法ではこの積分を一様分布に従う独立同分布の列 $U_0,U_1,\ldots$ を用いて $n^{-1}\sum_{i=0}^{n-1}(U_i+U_i^2)$ によって近似する．たとえば $n=100$ のとき，独立同分布 $(U_i;i=0,2,\ldots,99)$ の実現値に依存して，近似は 0.877423 や

[*1] 153-8914 東京都目黒区駒場 3-8-1

[*2] 本研究は科学研究費補助金若手研究 (B) (22740055) でサポートされた.

0.8048439 など様々な値を返すだろう．しかし，大数の法則により n が十分大きいなら真の値に近い事が保証される．さらに近似誤差は中心極限定理を用いて調べることも出来る．ただし，非決定論的であるから，まれに真値とかけ離れた値を返す可能性が常に存在することには注意すべきである．

以上がモンテカルロ法の考え方である．MCMC 法はモンテカルロ法の一つであるが，文字通りマルコフ連鎖を用いた分布によって実現値 $(x_i; i \geq 0)$ を生成する手法である．ただし，マルコフ連鎖とは確率空間 $(\Omega, \mathcal{F}, P)$ 上の確率変数列 $(X_n; n \geq 0)$ であり，$X_0, \ldots, X_{n-1}$ が得られた下での X_n の分布が X_{n-1} にのみ依存するものである．詳細は [30, 21] など参照．基本的なモンテカルロ法は独立同分布からの実現値を元にするのに対し，MCMC 法はマルコフ連鎖の分布を許すのである．一見些細な違いに思われるかもしれないが，この差によって多くの積分計算が可能になる．後で具体例を幾つか紹介する．

本論文ではマルコフ連鎖の中でも，独立型メトロポリス-ヘイスティングス (MH) 法と呼ばれる手法に焦点を当て，解析手法を紹介することが目的である．具体的な MCMC 法の構成は扱わないため，教科書 [25] を参照して欲しい．日本語の書籍でも MCMC 法の応用に関する著書は多いが，収束論などはそれらと比較するとまだ少ないが，有用な書籍としては例えば [34] がある．本論文が良書との架け橋になれば幸いである．残念ながら本論文では独立型 MH 法しか扱えなかったが，MCMC 法のその他の主な手法，ランダムウォーク型 MH 法やギブスサンプラーなどの手法についてや，その他応用に関する話題については教科書 [25] を参考にすると良い．モンテカルロ法や MCMC 法のベイズ統計の立場から見た歴史については [26] に詳しい．また，Diaconis らはより具体的なモデルに対して詳細な解析を行っており，本論文で紹介するアプローチと大きく異なる．彼らの数々の結果や，その研究視点について，レビュー [9] も面白い．

2 様々な手法の紹介

本節では独立型メトロポリス-ヘイスティングス (MH) 法に関連する手法を幾つか紹介する．

2.1 基本的なモンテカルロ法

ここでは，既に概要で述べた最も単純な形のモンテカルロ法についてより詳しく説明する．確率空間 $(\Omega, \mathcal{F}, \mathbf{P})$ 上に，E 値確率変数列 $(X_n; n \geq 0)$ があって，独立同分布であり，$\mathbf{P}(X_n \in A) = P(A)$ $(A \in \mathcal{E})$ であるとする．大数の法則により，$P(|f|) := \int_{x \in E} |f(x)| P(dx) < \infty$ であれば，$P_n(f)$ は $P(f)$ に概収束する．さらに $P(|f|^2) < \infty$ であれば中心極限定理

$$n^{1/2}(P_n(f) - P(f)) \Rightarrow N(0, \sigma^2)$$

が成立する．ただし，$\sigma^2 = P((f - P(f))^2)$ である．この事実から，$P_n(f)$ の近似の二乗誤差 $(P_n(f) - P(f))^2$ は $O_{\mathbf{P}}(n^{-1})$ のレートであることがわかる．よって n が十分大きければ近似は極めて $P(f)$ に近くなるだろう．収束理論も明快であり，プログラムも容易であるが，この近似手法には誤差が大きくなる因子が幾つかあり，注意して実行すべきである．

まず分散 $\sigma^2 < \infty$ であれば中心極限定理が成立するが，アルゴリズムの振る舞いは σ^2 の大きさに大きく依存する．誤差を小さくするためには n を大きく取らねばならないが，n の大きさは計算機の負担に対応するため，効率の面から問題がある．また，そもそも P に従う独立確率変数列を生成できなければ計算ができない．幸いなことにこれらの問題に対して様々な改善手法が存在する．以下で紹介する棄却法，importance sampling, MCMC 法はどれも改善手法であるが，状況に応じて適した改善手法を選ぶ必要がある．

2.2 棄却法

基本的なモンテカルロ法で積分近似が困難な場合に，しばしば用いられる方法が棄却法である．可測空間 $(E,\mathcal{E})$ 上の確率分布 Q で，$P \ll Q$，すなわち $A \in \mathcal{E}$ で，$Q(A)=0$ なら $P(A)=0$ となるものがあるとする．さらに

$$\frac{dP}{dQ}(x) \leq M \ (x \in E) \tag{1}$$

が満たされているとする．以後も出てくる記号 $dP/dQ(x)$ は P と Q の確率密度関数を p と q としたとき，$p(x)/q(x)$ の事である．確率空間 $(\Omega,\mathcal{F},\mathbf{P})$ 上に，E 値確率変数列 $(X_n;n\geq 0)$ があって，独立同分布であり，$\mathbf{P}(X_n \in A)=Q(A)$ $(A\in\mathcal{E})$ であるとする．また，$(X_n;n\geq 0)$ とも独立な独立確率変数列 $(U_n;n\geq 0)$ があって，一様分布に従うとする．このとき，$(X_n,U_n;n\geq 0)$ をもちいて，分布 P に従う独立確率変数列 $(Y_n;n\geq 0)$ を構成できる．まず $\tau_{-1}=-1$ とし，$i\geq 0$ について

$$\tau_i = \inf\{j>\tau_{i-1}; U_j \leq M^{-1}\frac{dP}{dQ}(X_i)\}$$

と定める．つぎに $Y_i = X_{\tau_i}$ とする．すると $(Y_n;n\geq 0)$ が求める確率変数列になっている．これをチェックするのはやさしい．たとえば Y_1 の分布が P であることを見てみよう．一般の場合については [8] 定理 2.3.1 を参照．まず

$$\mathbf{P}(Y_1\in A) = \sum_{i=1}^{\infty} P(Y_1 \in A, \tau_1 = i)$$

であるが，右辺の集合 $\{Y_1\in A,\tau_1=i\}$ は

$$\{X_i\in A, U_i \leq M^{-1}\frac{dP}{dQ}(X_i)\} \cap \bigcap_{j=1}^{i-1}\{U_j > M^{-1}\frac{dP}{dQ}(X_j)\}$$

となる．ここで任意の $A\in\mathcal{E}$ について

$$\mathbf{P}(X_i\in A, U_i\leq M^{-1}\frac{dP}{dQ}(X_i)) = \int_{\mathcal{X}} 1_A(x) M^{-1}\frac{dP}{dQ}(x)Q(dx) = M^{-1}P(A)$$

であり，独立性より $P(Y_1\in A,\tau_1=i)$ は

$$M^{-1}P(A)(1-M^{-1})^{i-1}$$

であるから，i について和をとれば $\mathbf{P}(Y_1\in A)=P(A)$ である．確率変数列 $(Y_n;n\geq 0)$ を計算機上で擬似的に生成する方法を棄却法という．応用上では，$dP/dQ\leq M$ なる M が存在してかつ既知であることは大きな制約であるが，分布 P と Q の関係が詳しく分かっている場合は有効な手法であり，例えば [8] の第 2.3 章では様々な棄却法が紹介されている．棄却法のパフォーマンスは上記の式展開で確認できるように，M の大きさに依存していて，小さいほうが有効である．

2.3 Importance Sampling

棄却法と同様に $P\ll Q$ なる確率分布 Q があり，確率空間 $(\Omega,\mathcal{F},\mathbf{P})$ および確率変数列 $(X_n,U_n;n\geq 0)$ は上と同じものとする．ただし (1) は仮定しない．すると

$$P(f) = Q(f\frac{dP}{dQ}) = \int_{x\in E} f(x)\frac{dP}{dQ}(x)Q(dx)$$

であるから，大数の法則より

$$Q_n(f\frac{dP}{dQ}) = n^{-1}\sum_{i=0}^{n-1} f(X_i)\frac{dP}{dQ}(X_i) \to P(f) \text{ a.s } (n\to\infty)$$

である．よって (1) 無しに，棄却法と同じ状況で積分の近似が出来る．ただし，P に従う独立同分布は構成していないことに注意する．応用上は Q の選択は有効性を左右する大きな問題である．ここでは確率分布 Q の有効性を図る指標として，$Q_n(f\frac{dP}{dQ})$ の分散を考える．場合によって $P_n(f)$ より $Q_n(f\frac{dP}{dQ})$ の方が近似誤差が小さいこともあることに注意する．分散は

$$\sigma_Q^2 := \int f(x)^2(\frac{dP}{dQ}(x))^2 Q(dx) - I^2.$$

の n^{-1} 倍で与えられる．シュワルツの不等式を用いれば，

$$\sigma_Q^2 \geq (\int |f(x)|P(dx))^2 - I^2$$

であり，等号は $Q(dx) = Q_f(dx) := |f(x)|P(dx)(\int |f(x)|P(dx))^{-1}$ で成立する．よって，分散を最小にするという意味で $Q = Q_f$ において最も有効である．しかし実際に計算機上で Q_f に従う独立同分布が生成できるとは限らない．予め独立同分布が生成できるようなクラス $\mathcal{Q}$ を考えて，そのなかで Q を選ぶ様々な方法がある [7].

3 マルコフ連鎖モンテカルロ法

前節では $P(f)$ の近似計算法を幾つか紹介した．しかし P の特性によっては棄却法も Importance sampling 法も困難なことがある．ベイズ統計学や統計物理学などで，しばしば P は分からないが，ある定数倍 cP は計算ができるという状況がある．このような構造がある場合に，マルコフ連鎖モンテカルロ（MCMC）法は有効である．ここでは簡単に定義を振り返るのみで，マルコフ連鎖自体の詳細の解説については [25] を参照.

例 1（イジングモデル） 実軸上の点 $i = 1,\ldots,n$ に観測 $s(i) = +1$ or -1 が定められており，$s = (s(i); i = 1\ldots,n)$ とする．このとき $H(s) = 2^{-1}\sum_{|i-j|=1} s(i)s(j)$ とし

$$P(\{s\}) = \frac{\exp(-H(s)/T)}{\sum_{t\in\{-1,1\}^{n^2}} \exp(-H(t)/T)} \tag{2}$$

によって $E = \{-1,+1\}^n$ 上の確率分布 P を定義する．すると，分母の計算は困難だが，$P(\{s\})$ の代わりに $\tilde{P}(\{s\}) = \exp(-H(s)/T)$ は計算が容易である．この問題には MCMC 法の中でもランダムウォーク型メトロポリス-ヘイスティングス (MH) 法が適している.

例 2（混合モデル） 確率変数 X の分布が

$$\theta\phi(x,0,1) + (1-\theta)\phi(x,1,1)$$

と表されているとする．ただし $\phi(x,\mu,1) = (2\pi)^{-1/2}\exp(-(x-\mu)^2/2)$ である．上の分布からの観測 $x_1,\ldots,x_{100}$ が得られているとする．一様事前分布のもと，θ の事後分布は

$$p(\theta) = Z^{-1}\prod_{i=1}^{100}(\theta\phi(x_i,0,1) + (1-\theta)\phi(x_i,1,1))$$

となる．ただし Z は $\int p(\theta)d\theta = 1$ を満たすように取られた正規化定数である．この場合も $p(\theta)$ は計算が難しいが，正規化定数を除いた $\tilde{p}(\theta)$ は計算できる．この問題には MCMC 法の中でも，ギブスサンプラーが自然に定義できるが，独立型 MH 法が有効である．

3.1 マルコフ連鎖モンテカルロ法の概要

3.1.1 独立型メトロポリス-ヘイスティングス法

Importance sampling 法の場合と同じように，確率空間 $(\Omega, \mathcal{F}, \mathbf{P})$ 上に独立確率変数列 $(X_n, U_n; n \geq 0)$ があって，X_n の分布は Q，U_n は一様分布に従うとする．確率分布 Q は P について絶対連続とする．このとき $(Y_n; n \geq 0)$ を以下のように定める．まず $Y_0 = X_0$ とし，逐次的に $i \geq 0$ で

$$Y_{i+1} = \begin{cases} X_{i+1} & \text{if } U_i \leq \frac{dP}{dQ}(X_{i+1})/\frac{dP}{dQ}(X_i) \\ X_i & \text{otherwise} \end{cases}$$

とする．すると $(Y_n; n \geq 0)$ は独立同分布ではなく，マルコフ連鎖になる．ここで，上の Y_{i+1} の定義において，P を $\tilde{P} = cP$ で置き換えても Y の分布は変わらないことに注意する．マルコフ連鎖は独立同分布に従う確率変数列と近い性質を持ち，$P(|f|) < \infty$ であるなら

$$K_n(f) := n^{-1}\sum_{i=1}^{n} f(Y_i)$$

とおくと $K_n(f) \to P(f)$ であり，あとで示すように，中心極限定理も成立する．

マルコフ連鎖の遷移確率カーネル K は具体的には

$$K(x, dy) = \alpha(x, y)Q(dy) + \big(1 - \int_{z \in E} \alpha(x, z)Q(dz)\big)\delta_x(dy)$$

と書ける．ここで $\alpha(x, y)$ は $W(x) = dP/dQ(x)$ として

$$\alpha(x, y) \leq \alpha^*(x, y) := \min\{1, W(y)/W(x)\}$$

かつ $\alpha(x, y)W(x)$ が x, y について対称になる関数である．ただし，$\alpha = \alpha^*$ とすることが漸近分散の意味で最適である ([24, 32])．確率変数列 $(Y_n; n \geq 0)$ を擬似的に生成するアルゴリズムを，（独立型）メトロポリス-ヘイスティングス (MH) 法と呼ぶ．

3.1.2 ランダムウォーク型メトロポリス-ヘイスティングス法

状態空間 E は有限集合 $E = \{1, 2, \ldots, m\}$ とする．このとき先程の独立型 MH 法と異なって，X_n, Y_n は逐次的に生成する．既に $Y_1, \ldots, Y_i$ が得られているとき，Z_{i+1} を -1 と $+1$ を等確率で取る確率変数として

$$X_{i+1} = Y_i + Z_{i+1}$$

とする．$\alpha(x, y) = \min\{1, P(\{y\})/P(\{x\})\}$ として，確率 $\alpha(Y_i, X_{i+1})$ で $Y_{i+1} = X_{i+1}$，残りの確率で $Y_{i+1} = Y_i$ となるように Y_{i+1} を定める．ただし $Y_i = 1$ かつ $Z_{i+1} = -1$，もしくは $Y_i = m$ かつ $Z_{i+1} = +1$ のときは $Y_{i+1} = Y_i$ となるものとする．するとやはり出来上がった $(Y_i; i \geq 0)$ はマルコフ連鎖であり，例えば $P(\{x\}) > 0$ $(x = 1, \ldots, m)$ であれば $K_n(f) \to P(f)$ が成立するし，中心極限定理が成り立つことがわかる [30]．

イジングモデルでは，$\{-1,+1\}$ の系列 $E=\{-1,+1\}^n$ が状態空間であり，$Y_n=s\in E$ のとき，Y_{n+1} は X_n と一カ所だけ異なる状態への一様分布とする．もしくは，$Y_n=(s(1),\ldots,s(n))$ のベクトルの一カ所を一様に選び，X_{n+1} は Y_n とその一カ所のベクトルだけ異なる物とする．先ほどと同じように X_{n+1} は確率 $\alpha(Y_n,X_{n+1})=\min\{1,P(\{X_{n+1}\})/P(\{Y_n\})\}$ で採択される．ここでは Z_{i+1} を明示的に書いていないが，自然に定義されている事に注意する．

状態空間が $\mathbf{R}^p$ のときは，Z_{i+1} の分布は正規分布 $N(0,\sigma^2 I)$ とする．ただし I は $p\times p$ の単位行列である．確率分布 P の密度を $p(x)$ と書くと，$\alpha(x,y)=\min\{1,p(y)/p(x)\}$ として上と全く同じように Y_i を Y_{i+1} へ更新する．離散の場合と異なり，極限定理については注意すべきであるが，本論文ではあまり扱わない．第4章を参照．

3.1.3 ギブスサンプラー

ギブスサンプラーでは，確率分布 $P(dx)$ は，より広い状態空間 $E\times F$ 上の分布 $P^*(dx,dy)$ の E への制限としてとらえる．$P(dx)=P^*(dx)=P^*(dx\times F)$ および $P^*(dy)=P^*(E\times dy)$ とし，

$$P^*(dx,dy)=P^*(dx)P^*(dy|x)=P^*(dy)P^*(dx|y)$$

となる $P^*(dy|x)$，$P^*(dx|y)$ があるとする．このとき $Y_0,\ldots,Y_i$ が得られているとき，

$$X_{i+1}\sim P^*(dx|y=Y_i),\ Y_{i+1}\sim P^*(dy|x=X_{i+1})$$

として $(Y_i;i\geq 0)$ を定義していく．独立型 MH 法やランダムウォーク型 MH 法と構造が大きく異なるが，より広い枠組みの MH 法の一つとして捉えられる．極限定理については第4章を参照．

先ほどの例2は $E=\Theta=[0,1]$ および x の代わりに θ という記号を用いている事に注意すれば，P^* は θ の空間 $E=[0,1]$ と $F=\{0,1\}^{100}$ の直積上の分布として次のように定義される：

$$P^*(d\theta,d(y_1,\ldots,y_{100}))=Z^{-1}\theta^{\sum_{i=1}^{100}y_i}(1-\theta)^{\sum_{i=1}^{100}(1-y_i)}\prod_{i=1}^{100}\phi(x_i,y_i,1).$$

上記の分布 $P^*(d\theta,dy)$ を y について積分すれば $P^*(d\theta\times F)=P(d\theta)$ となることに注意する．このとき $P^*(dy|\theta)$ はパラメータ $\alpha=1+\sum_{i=1}^n y_i,\beta=1+\sum_{i=1}^n(1-y_i)$ のベータ分布 $\mathrm{Beta}(\alpha,\beta)$ であり，$P^*(d\theta|y)$ は正規分布として書ける．

これから独立型 MH 法の K について詳しく解析し，$K_n(f)$ について考察する．

3.2 エルゴード性について

前節で $K_n(f)$ は n が十分大きければ正しく $P(f)$ を近似できることに言及した．このような近似が正当化されるためには，遷移確率カーネル K が，幾つかの良い性質を持つことが必要である．具体的には規約性，aperiodicity，再帰性およびハリス再帰性である．これらの性質を満たすマルコフ連鎖はエルゴード性を持つといい，その場合 $K_n(f)\to P(f)$ となる．詳しい用語の定義については適当な書物 [20] を参照して欲しいが，おおよそ次のような意味である．以下で用いる遷移確率カーネル K^i は次で定義される：

$$K^1(x,dy)=K(x,dy),K^{i+1}(x,dz)=\int_{y\in\mathcal{E}}K^i(x,dy)K(y,dz)\ (i\geq 1).$$

規約性 任意の出発点 $X_0=x\in E$ から，任意の**ある程度おおきな**集合 $A\in\mathcal{E}$ に移りうる，すなわち，$P_x(\bigcup_{n=1}^\infty\{X_n\in A\})>0$ である．

apperiodicity 集合 E の分割 E_1, E_2 で $K(x,E_2)=1$ $(x\in E_1)$, $K(x,E_1)=1$ $(x\in E_2)$ となるとき，$(E_i;i=1,2)$ を 2-サイクルという．一般に d-サイクルがあると，K^{dn+i} $(i=1,2,\ldots,d-1)$ の性質が i によって大きく異なり $\lim_{n\to\infty}K^n$ は考えにくい．サイクルのない遷移確率カーネルを aperiodic という．

再帰性 任意の $x\in E$ から出発したマルコフ連鎖は，任意のある程度おおきな集合 $A\in\mathcal{E}$ に何回でも訪れうる，すなわち $\sum_{i=1}^n K^n(x,A)=+\infty$ $(x\in A)$ となる．

ハリス再帰性 任意の $x\in E$ から出発したマルコフ連鎖は，任意のある程度おおきな集合 $A\in\mathcal{E}$ に必ず無限回訪れる．

これらの性質のもと，$K_n(f)$ の $P(f)$ への収束や $\|K^n(x,\cdot)-P\|=2^{-1}\sup_{A\in\mathcal{E}}|K^n(x,A)-P(A)|$ の収束が成り立ち，さらにその収束レートについてもマルコフ連鎖のエルゴード性の理論は有用な情報を与える．ドリフト関数を用いたエルゴード性の必要十分条件 [33, 23] や，スプリッティング技術 [22, 4] など，後に MCMC 法で重要になるマルコフ連鎖の理論が 1970, 80 年代にまとまり，これらの手法を用いて，R. L. Tweedie や G. O. Roberts，そして J. Rosenthal といった研究者が中心になり，2000 年代に入るまでには MCMC 法のエルゴード性の理論は整備された．まとまったレビューとして [26] やより理論的レビュー [28] がある．Tierney による [31] はマルコフ連鎖のエルゴード性を MCMC に適用する鮮やかなレビューである．独立型 MH 法に対するエルゴード性の議論は容易である．$P\ll Q$ であれば独立型 MH 法はエルゴード性を持ち [31] したがって任意の P 可積分関数 f について

$$K_n(f)\to P(f)$$

なる概収束が成立する．このため仮定 $P\ll Q$ という極めて一般的な条件のもとで，独立型 MH 法は正しく $P(f)$ を近似できるのである．また (1) なる有界な M が存在することと，より良いエルゴード性である一様エルゴード性が成り立つことは同値である [19]．この条件はマルコフ連鎖の理論では Doeblin 条件と呼ばれ，直ちに $f\in L^2(P)$ について

$$n^{-1/2}\left(\sum_{i=1}^{n}K_n(f)-P(f)\right)$$

が漸近正規性を持つ事を示せる．中心極限定理の詳細は次節で扱う．また，同じ仮定のもと Doeblin 条件から独立型 MH 法において $\|K^n(x,\cdot)-P\|$ は $x\in E$ の値に一様に指数オーダーで収束する．このように，独立型 MH 法の解析は容易に展開できて様々な大域的性質を導くが，(1) を満たさない場合の評価や，満たしたとしても，より詳細の振る舞いを記述するのは難しい．しかし独立型 MH 法に限れば，マルコフ連鎖のエルゴード性の議論を通さず，より古典的な手法で精密な評価が可能である．

3.3 MCMC の中心極限定理

ここでは $P(f)$ の近似 $K_n(f)$ の誤差が $n^{-1/2}$ のオーダーであることを，中心極限定理の成立を示すことによってみてみる．とくに (1) という強い仮定のもとでは中心極限定理の成立は容易に示せるが，より一般的な仮定のもとで考える．まず Y がエルゴード性をもつことを仮定する．この仮定をチェックすることは独立型 MH 法であれば $P\ll Q$ をチェックすれば十分であった．また簡単のため Y_0 の分布が P すなわち，Y が定常であることを仮定する．この仮定は実用上ナンセンスであるが，以下の議論は Y がエルゴード性を持つ限り，一般の Y_0 で成り立つ．ここでは関数 f を $L_0^2(P)=\{f:E\to\mathbf{R};\int f(x)P(dx)=0,\int f(x)^2P(dx)<\infty\}$ の元と考えると都合が良い．内積は $(f,g)=\int P(dx)f(x)g(x)$ と書く．遷移確率カーネル K は $f(x)$ を

$Kf(x) = \int_{y \in E} K(x, dy) f(y)$ とするマップと捉えると，縮小作用素である．独立型 MH 法に限らず，ほとんど全ての MCMC 法は

$$P(dx)K(x, dy) = P(dy)K(y, dx) \tag{3}$$

となる遷移確率カーネルを定める．この条件を満たすマルコフ連鎖はリバーシブルであるという．リバーシブルであることは，K が自己共役作用素であることを意味するから，ある単位の分解 $(E(\Lambda); \Lambda \in \mathcal{B}[-1, 1])$ があって，

$$K = \int \lambda dE(\lambda)$$

と書ける．状態空間が有限集合の場合，上の関係式は K の対角化に対応する [24]．リバーシブルなマルコフ連鎖に対する中心極限定理の一般的結論が [17] によって得られた．

仮に f が $(I - K)^{-1}$ の領域に入っているとし，$g = (I - K)^{-1} f$ とする．すると

$$K_n(f) = n^{-1} \sum_{i=1}^{n} (I - K) g(Y_i) = n^{-1} \sum_{i=1}^{n} (g(Y_i) - Kg(Y_{i-1})) + n^{-1}(-Kg(Y_n) + Kg(Y_0))$$

となる．適当な条件のもと右辺第二項が無視できるとして，右辺の第一項の振る舞いを考えると，これは $\mathcal{F}_n := \sigma(Y_0, \ldots, Y_n)$-マルチンゲールであるから，以下の右辺が存在する限り

$$n^{-1} \sum_{i=1}^{n} E[(g(Y_i) - Kg(Y_{i-1}))^2 | \mathcal{F}_{i-1}] \to E[(g(Y_1) - Kg(Y_0))^2]$$

なる収束が成り立ち，さらに右辺は，K の対称性に注意して $\|Kg\|^2 = \int P(dx)K(x, dy)K(x, dz)g(y)g(z) = \int P(dy)K(y, dx)K(x, dz)g(y)g(z) = (g, K^2 g)$ であることを用いると

$$\|g\|^2 - \|Kg\|^2 = (g, (I - K^2)g) = \int \frac{1 - \lambda^2}{(1 - \lambda)^2} d\|E(\lambda)f\|^2 = \int \frac{1 + \lambda}{1 - \lambda} d\|E(\lambda)f\|^2$$

となる．従って

$$\int \frac{1 + \lambda}{1 - \lambda} d\|E(\lambda)f\|^2 < \infty \tag{4}$$

であれば，仮に f が $(I - K)^{-1}$ の領域に入っているなら，マルチンゲール差分の中心極限定理（例えば [12]）から，$n^{1/2}(K_n(f) - P(f))$ が正規分布に収束する，とくに $n^{1/2}(K_n(f) - P(f)) = O_P(1)$ であることがわかる．実際には (4) が成立しても f は必ずしも $(I - K)^{-1}$ の領域に入っていないが，任意の $\epsilon > 0$ について $(I + \epsilon - K)^{-1}$ の領域には入っており，あとは $\epsilon \to 0$ として $(I - K)^{-1}$ に入っていない f についても上の議論が正当化される [17]．よってエルゴード性をもつ MCMC 法であれば (4) を示せれば中心極限定理が従う．ただし，一般に (4) の左辺の値を知ることは難しい．

中心極限定理に関して，前述のマルコフ連鎖のエルゴード性からのアプローチでの結果を付記しておく．一般に幾何型エルゴード性のときは [5, 27] により $f \in L^2(P)$ であれば，また多項式型収束なら [15] により，ある $\alpha > 0$ があって $f \in L^{2+\alpha}(P)$ であれば (4) が成り立つことが示されている．幾何エルゴード性や多項式型エルゴード性は，ドリフト条件の構成によってチェックすることができて（例えば [19, 29]），かつ中心極限定理が成立した場合にその漸近分散 (4) の左辺がドリフト関数によって評価ができる．そのため，通常は実用的理由で，スペクトル分解を用いた議論よりこちらのマルコフ連鎖のエルゴード性を用いた議論が好まれる．しかし，次節で見るように，独立型 MH 法にたいしてはドリフト関数を用意して評価するまでもなく，解析的に漸近分散 (4) が求まる．

3.4 スペクトル分解

ここでは，具体的に独立型 MH 法で定められる遷移確率カーネル K のスペクトル分解を行う．状態空間 E が有限である場合は [18] によってスペクトル分解が得られている．状態空間が一般の場合もほとんど既知の事実であるが，現在投稿準備中の論文に記載予定である．

ここで幾つか仮定を置く．まず $\alpha = \alpha^*$，すなわち独立型 MH 法は最良の α を用いるものとする．また $W(x) = dP/dQ(x)$ とし，W が分布 Q についてアトムを持たない，すなわち $Q(W = c) > 0$ なる $c \geq 0$ が存在しないことを仮定する．このとき作用素 U, V を $\gamma(x,y) = 1_{\{W(x)>W(y)\}}$ を用いて

$$Uf(x) = (\int_{y\in E} P(dy)\gamma(x,y)(f(x) - f(y)))(\int_{y\in E} P(dy)\gamma(x,y))^{-2}$$

$$Vf(x) = \int_{y\in E} P(dy)\gamma(x,y)(f(x) + f(y))$$

とする．簡単な計算により，以下の補題が得られる．

補題 1 上の仮定のもと，$UV = VU = I$.

さらに，棄却率 $\lambda(x) = 1 - \int_{x\in E} \beta(x,z)P(dz)$ とし，$\Lambda f(x) = \lambda(x)f(x)$ とおく．また，作用素の族 $(E(A); A \in \mathcal{E})$ を

$$E(A)f(x) = \int_{y\in A, z\in E} V(x,dy)U(y,dz)f(z)$$

により定める．ここで E は，Λ と異なり，P を $\tilde{P} = cP$ に置き換えても良い，すなわち．定数倍を除いてわからなくても求まることに注意する．すると $(E(A); A \in \mathcal{E})$ は単位の分解であり，次を得る．

命題 1 補題と同じ仮定のもと $K = \int_{x\in E} \lambda(x) dE(x)$.

よって，複雑であってここには記載しないが，具体的に漸近分散などの統計量が表現出来る．これを用いて，計算の容易な収束レートの評価の導出ができれば Q の選択に実用的に使えるだろう．このようなスペクトル分解はランダムウォーク型 MH 法などでは得られておらず，独立型 MH 法のシンプルさの特徴である．

4 その他のマルコフ連鎖モンテカルロ法についての補足

独立型 MH 法は解析的にシンプルであったため，具体的なスペクトル分解によって漸近分散などの評価が得られた．実用上は様々な MCMC 法が使われており，それらについては正確な漸近分散などの導出は難しく，異なったアプローチによって振る舞いの評価をする必要がある．状態空間 E が有限の時は P に応じて様々な評価が知られている．一方，E が有限集合と限らないときは，フォスター-リヤプノフ型のドリフト条件

$$PV \leq cV + b1_C$$

を用いた評価が一般的である．ただし，$V : E \to [1, \infty]$ は可測関数で，$c, b > 0$ は定数，C はコンパクト集合である．このようなドリフト条件の存在の具体的構成が一次元の E の場合 [19] および多次元の場合 [29] や [13] によって，また必要条件についても [14] によって調べられた．

ギブスサンプラーはベイズ統計において基本的ツールになっている．この場合もスペクトル分解の具体的表現は難しく，さらにドリフト条件での評価は有効でないことが多い．しかし局所漸近正規性など大標本理論を用いた評価が可能であり，一般的な条件のもとで有効性の概念が定義できる [16].

近年では MCMC 法のエルゴード性の解析の議論は落ち着き，より発展的なアルゴリズムの解析が研究されている．MCMC 法では $(X_i; i \geq 0)$ は常にマルコフ連鎖であったが，適合的マルコフ連鎖モンテカルロ (AMCMC) 法 [11] では，マルコフ連鎖の仮定をやや緩めることを目的としている．解析はマルコフ連鎖と比べ極めて困難であるが，限定的でありながら，AMCMC 法の様々なエルゴード性の結果が得られている ([2, 3])．一方，逐次モンテカルロ (SMC) 法と MCMC 法の組み合わせのアルゴリズムも様々提案されており ([10, 1, 6])，実用的，理論的に活発な分野になっている．

本論文では独立型 MH 法に焦点を当てて解析法を紹介したが，手法や状況に応じて解析手法も様々であり，普遍的な最良な方法は存在しない．独立型 MH 法に限っても，本論文では regeneration によるアプローチを載せることはできなかったが，場合によっては視点を変えて解析したほうが都合が良いこともある．モデルの特性を把握し，有効な手法を選択すべきである．

参考文献

[1] Christophe Andrieu, Arnaud Doucet, and Roman Holenstein, *Particle Markov chain Monte Carlo methods*, Journal of the Royal Statistical Society. Series B (Methodological) **72** (2010), no. 3, 269–342.

[2] Christophe Andrieu and Eric Moulines, *On the ergodicity properties of some adaptive MCMC algorithms*, Ann. Appl. Probab **16** (2006), no. 3, 1462–1505.

[3] Yves Atchade and Gersende Fort, *Limit theorems for some adaptive MCMC algorithms with subgeometric kernels*, Bernoulli **16** (2010), no. 1, 116–154.

[4] K. B. Athreya and P. Ney, *A new approach to the limit theory of recurrent Markov chains*, Transactions of the American Mathematical Society **245** (1978).

[5] Kung Sik Chan and Charles J. Geyer, *Discussion: Markov Chains for Exploring Posterior Distributions*, Annals of Statistics **22** (1994), no. 4, 1747–1758.

[6] N. Chopin, P.E. Jacob, and O. Papaspilippoulos, *SMC2: A sequential Monte Carlo algorithm with particle Markov chain Monte Carlo updates* , Arxiv (2011).

[7] Pieter-Tjerk de Boer, Dirk P. Kroese, Shie Mannor, and Reuven Y. Rubinstein, *A tutorial on the cross-entropy method*, Annals of Operations Research **134** (2005), 19–67.

[8] Luc Devroye, *Non-Uniform Random Variate Generation*, Springer-Verlag, New York, 1986.

[9] Persi Diaconis, *The markov chain monte carlo revolution*, Bulletin of American Mathematical Society **46** (2008), 179-205.

[10] Walter R. Gilks and Carlo Berzuini, *Following a moving target - Monte Carlo inference for dynamic Bayesian models*, Journal of the Royal Statistical Society. Series B (Methodological) **63** (2001), no. 1.

[11] Heikki Haario, Eero Saksman, and Johanna Tamminen, *An adaptive Metropolis algorithm*, Bernoulli **7** (2001), no. 2, 223–242.

[12] Inge S. Helland, *Central Limit Theorems for Martingales with Discrete or Continuous Time*, Scandinavian Journal of Statistics **9** (1982), no. 2, 79–94.

[13] S. F. Jarner and E. Hansen, *Geometric ergodicity of Metropolis algorithms*, Stochastic Processes and their Applications **85** (2000), 341–361.

[14] S. F. Jarner and R. L. Tweedie, *Necessary conditions for geometric and polynomial ergodicity of random-walk-type Markov chains*, Bernoulli **9** (2003), no. 4, 559–578.

[15] Kengo Kamatani, *Metropolis-Hastings Algorithms with acceptance ratios of nearly 1*, Annals of the Institute of Statistical Mathematics **61** (2009), no. 4, 949–967.

[16] ———, *Weak Convergence of Markov Chain Monte Carlo Methods and its Application to Regular Gibbs Sampler*, Arxiv (2010).

[17] C. Kipnis and S. R. S. Varadhan, *Central limit theorem for additive functionals of reversible Markov processes and applications to simple exclusions*, Communications in Mathematical Physics **104** (1986), no. 1, 1–19.

[18] Jun S. Liu, *Metropolized independent sampling with comparisons to rejection sampling and importance sampling*, Statistics and Computing **6** (1996), 113–119.

[19] K. L. Mengersen and R. L. Tweedie, *Rates of convergence of the Hastings and Metropolis algorithms*, The Annals of Statistics **24** (1996), 101–121.

[20] S. P. Meyn and R. L. Tweedie, *Markov Chains and Stochastic Stability.*, Springer, 1993.

[21] J. R. Norris, *Markov Chains*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, 1998.

[22] Esa Nummelin, *A splitting technique for Harris recurrent Markov chains*, Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete **43** (1978), no. 309–318.

[23] Esa Nummelin and Pekka Tuominen, *Geometric ergodicity of Harris recurrent Markov chains with applications to renewal theory*, Stochastic Processes and their Applications **12** (1982), no. 187–202.

[24] P. H. Peskun, *Optimum Monte-Carlo sampling using Markov chains*, Biometrika **60** (1973), no. 3, 607–612.

[25] Christian P. Robert and George Casella, *Monte Carlo Statistical Methods.*, 3rd ed., Springer, 2004.

[26] ———, *A Short History of Markov Chain Monte Carlo*, Arxiv (2011).

[27] G. O. Roberts and J. S. Rosenthal, *Geometric ergodicity and hybrid Markov chains*, Electron. Comm. Probab **2** (1997), 13–25.

[28] ———, *General state space markov chains and mcmc algorithms*, Probability Surveys **1** (2004), 20–71.

[29] G. O. Roberts and R. L. Tweedie, *Geometric convergence and central limit theorems for multidimensional Hastings and Metropolis algorithms*, Biometrika **83** (1996), 95–110.

[30] Eugene Seneta, *Non-negative matrices and Markov chains*, Springer series in statistics, vol. 2, Birkhäuser, 2006.

[31] L. Tierney, *Markov Chains for Exploring Posterior Distributions (with discussion)*, The Annals of Statistics **22** (1994), no. 4, 1701–1762.

[32] ———, *A note on Metropolis-Hastings kernels for general state spaces*, The Annals of Applied Probability **8** (1998), no. 1, 1–9.

[33] Richard L. Tweedie, *Sufficient conditions for ergodicity and recurrence of Markov chains on a general state space*, Stochastic Processes and their Applications **3** (1975), no. 385–403.

[34] 伊庭 幸人, 種村 正美, 大森 裕浩, 和合 肇, 佐藤 整尚, and 高橋 明彦, **計算統計 *2* マルコフ連鎖モンテカルロ法とその周辺**, 岩波書店, 2005.