

# PRML 11.2 - 11.4

2022/09/13 esonomot

## 11.2 Markov Chain Monte Carlo

- これまで見た棄却サンプリングなどは、高次元になると sampling 効率が指数関数的に悪くなっていくという「次元の呪い」が存在していた
- そこで次元に依らない別の sampling 手法として、最終的に得られる数 (列) が欲しい分布  $p(z)$  に従うよう、適当な値  $z$  から始める何らかの連続的な sampling を考える
- この手法が満たして欲しい性質は以下の 2 つである
  1. 欲しい分布  $p(z)$  を不変に保つ sampling が存在する
  2. どんな分布から始めても sampling が最終的に欲しい分布  $p(z)$  に収束する
- **Markov chain** を使うとこの 2 つの要請を満たす条件を簡単に導くことができる

---

“ ここでも  $p(z) = \tilde{p}(z)/Z_p$  としたとき  $\tilde{p}(z)$  はすぐに計算できるとする

## 11.2.1 Markov Chains

- A **first-order Markov chain** is a series of random variables  $\mathbf{z}^{(1)}, \dots, \mathbf{z}^{(M)}$  following the conditional independence property

$$p(\mathbf{z}^{(m+1)} | \mathbf{z}^{(1)}, \dots, \mathbf{z}^{(m)}) = p(\mathbf{z}^{(m+1)} | \mathbf{z}^{(m)}),$$

for  $m \in \{1, \dots, M - 1\}$ .

- 従って、以下の 2 つがわかれば Markov chain を指定することができる
  - 初期値  $p(\mathbf{z}^{(0)})$
  - **遷移確率** (transition probability)  $T_m(\mathbf{z}^{(m+1)}, \mathbf{z}^{(m)}) \equiv p(\mathbf{z}^{(m+1)} | \mathbf{z}^{(m)})$
- 全ての  $m$  について遷移確率の同じ Markov chain は**均一マルコフ連鎖** (homogeneous Markov chain) と呼ばれる

# Invariance of Distributions

- Markov chain では、特定の変数の周辺分布はその一つ前の変数のみで表される

$$p(\mathbf{z}^{(m+1)}) = \sum_{\mathbf{z}^{(m)}} p(\mathbf{z}^{(m+1)} | \mathbf{z}^{(m)}) p(\mathbf{z}^{(m)}).$$

- 分布が Markov 連鎖の前後で変わらないとき、その分布は Markov 連鎖に関して **不変** (もしくは定常) と呼ばれる
- 例えば、遷移確率が  $T(\mathbf{z}', \mathbf{z})$  である均一マルコフ連鎖において、

$$p^*(\mathbf{z}) = \sum_{\mathbf{z}'} T(\mathbf{z}', \mathbf{z}) p^*(\mathbf{z}'),$$

が成り立つとき、分布  $p^*(\mathbf{z})$  は不変である

# Detailed Balance

- 1 つ目の要請である「不変分布の存在」を保証するのが**詳細釣り合い条件**である
- 求めたい分布が不変分布であることを保証する十分条件は、分布  $p^*(z)$  に対し

$$p^*(z)T(z, z') = p^*(z')T(z', z),$$

で定義される**詳細釣り合い条件** (detailed balance) が満たされるように遷移確率を選ぶことである

- 詳細釣り合い条件を満たすならばその分布が不変であることは容易に導かれる

$$\sum_{z'} p^*(z')T(z', z) = p^*(z) \sum_{z'} T(z, z') = p^*(z).$$

# Ergodicity

- 2 つ目の要請である「分布の収束性」を保証するのが**エルゴード性**である
- Markov chain におけるエルゴード性とは Markov chain が既約かつ非周期的であることである
  - 既約：Markov chain がこれ以上小さい Markov chain に分割できない
  - 非周期的：Markov chain に循環がない
- エルゴード性を持つマルコフ連鎖 (エルゴード的マルコフ連鎖) における不変分布は**平衡分布**と呼ばれ、平衡分布はただ 1 つだけ存在することが知られている

# Summary

- 任意の Markov chain が少なくとも一つの不変分布を持つことから
  - 求めたい分布が詳細釣り合い条件を満たすならば求めたい分布は不変分布である
  - エルゴード的 Markov chain は任意の初期分布に対して不変分布に収束するという性質が導かれる
- 結果として、**求めたい分布に関して詳細釣り合い条件を満たすようにエルゴード的 Markov chain を構成すれば、求めたい分布に収束するような数 (列) が得られる**ということがわかった

## 11.2.2 The Metropolis-Hastings algorithm

- 上記条件を満たす最も一般的な方法が Metropolis-Hastings アルゴリズムである
- **The Metropolis-Hastings algorithm**

At step  $\tau$  of the algorithm, in which the current state is  $z^{(\tau)}$ , we draw a sample  $z^*$  from the distribution  $q_k(z|z^{(\tau)})$  and then accept it with probability

$$A_k(z^*, z^{(\tau)}) = \min \left( 1, \frac{\tilde{p}(z^*)q_k(z^{(\tau)}|z^*)}{\tilde{p}(z^{(\tau)})q_k(z^*|z^{(\tau)})} \right),$$

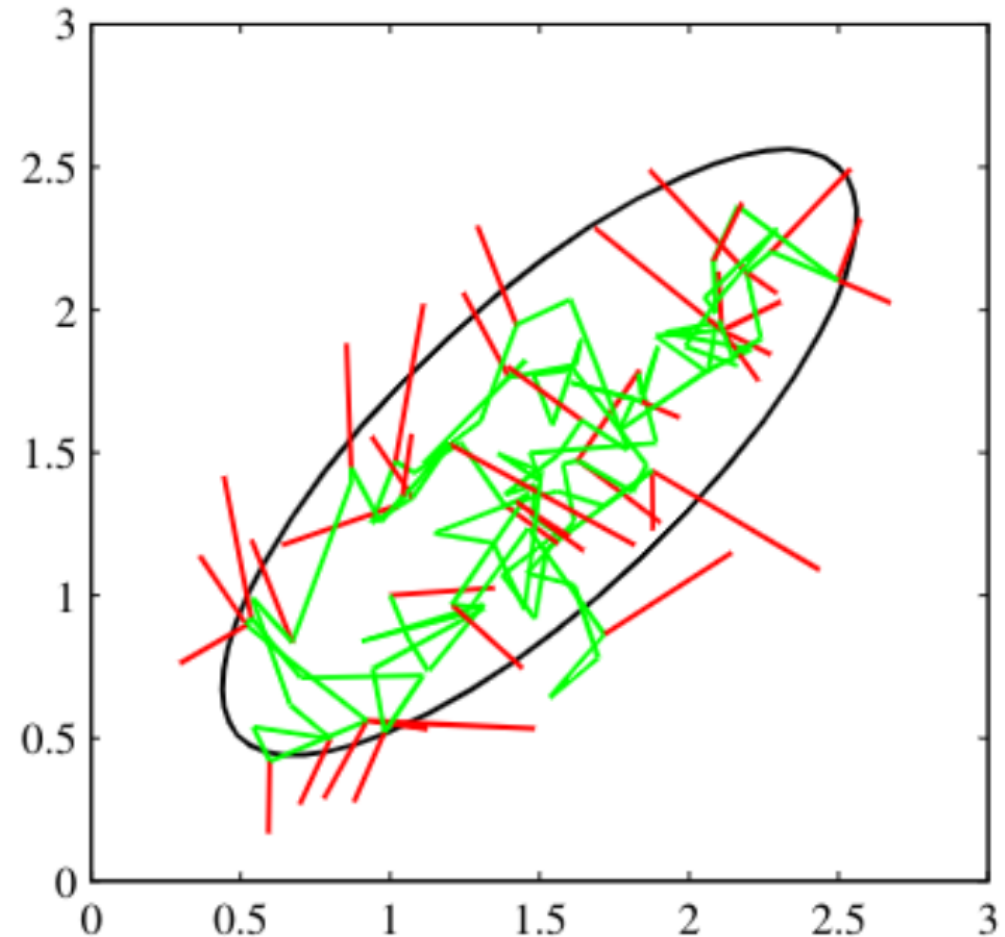
where  $k$  labels the members of the set of possible transitions being considered.

- 対称な提案分布の場合、 $q_k$  が分母分子でキャンセルされ  $\tilde{p}$  の大小のみで棄却が決まる **Metropolis algorithm** となる



# Metropolis Algorithm

**Figure 11.9** A simple illustration using Metropolis algorithm to sample from a Gaussian distribution whose one standard-deviation contour is shown by the ellipse. The proposal distribution is an isotropic Gaussian distribution whose standard deviation is 0.2. Steps that are accepted are shown as green lines, and rejected steps are shown in red. A total of 150 candidate samples are generated, of which 43 are rejected.



# Detailed Balance

- Metropolis-Hasting algorithm では、棄却条件は

$$\begin{aligned} p(z)q_k(z'|z)A_k(z', z) &= \min(p(z)q_k(z'|z), p(z')q_k(z|z')), \\ &= \min(p(z')q_k(z|z'), p(z)q_k(z'|z)), \\ &= p(z')q_k(z|z')A_k(z, z'). \end{aligned}$$

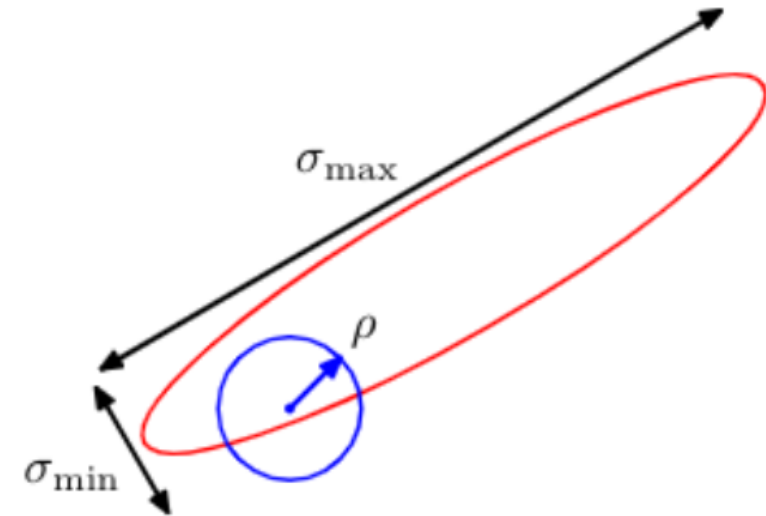
と表されるので、詳細釣り合い条件が満たされていることがわかる

# Choice of Proposal Distribution

- 提案分布の選択はアルゴリズムの性能に大きな影響を与える
  - 小さな分散 → 棄却率は低いが状態空間内の進みは遅い
  - 大きな分散 → 棄却率は高いが状態空間内の進みは早い
- $z$  の要素間に強い相関を持つ多変量 (2変量) 分布  $p(z)$  を考える (Figure 11.10)
- このとき、
  - 高い棄却率を招かない最大の分散  $\rho$  のスケールは  $\sigma_{\min}$
  - 分散が大きな方向に沿ってランダムウォーク的に探索することから、独立なサンプルは  $(\sigma_{\max}/\sigma_{\min})^2$  程度のステップで得られる

# Illustration of Exploration

**Figure 11.10** Schematic illustration of the use of an isotropic Gaussian proposal distribution (blue circle) to sample from a correlated multivariate Gaussian distribution (red ellipse) having very different standard deviations in different directions, using the Metropolis-Hastings algorithm. In order to keep the rejection rate low, the scale  $\rho$  of the proposal distribution should be on the order of the smallest standard deviation  $\sigma_{\min}$ , which leads to random walk behaviour in which the number of steps separating states that are approximately independent is of order  $(\sigma_{\max}/\sigma_{\min})^2$  where  $\sigma_{\max}$  is the largest standard deviation.



# Choice of Proposal Distribution in General

- 実は 2 次元の場合、棄却率の高さと探索範囲の広さが相殺されるため、 $\rho$  を  $\sigma_{\max}, \sigma_{\min}$  のどちらに設定しても探索効率は変わらない
- そのため、より一般に、多変量ガウス分布では 2 番目に小さな標準偏差  $\sigma_2$  を用いて探索するのが効率がよく、その結果独立なサンプルを得られるステップはおよそ  $(\sigma_{\max}/\sigma_2)^2$  となる

## 11.3 Gibbs Sampling

- Gibbs sampling は先の Metropolis-Hastings algorithm の特別な場合に相当する

### Gibbs Sampling

1. Initialize  $\{z_i : i = 1, \dots, M\}$
2. For  $\tau = 1, \dots, T$ :
  - Sample  $z_1^{(\tau+1)} \sim p(z_1 | z_2^{(\tau)}, z_3^{(\tau)}, \dots, z_M^{(\tau)})$ .
  - Sample  $z_2^{(\tau+1)} \sim p(z_2 | z_1^{(\tau+1)}, z_3^{(\tau)}, \dots, z_M^{(\tau)})$ .
  - $\vdots$
  - Sample  $z_j^{(\tau+1)} \sim p(z_j | z_1^{(\tau+1)}, \dots, z_{j-1}^{(\tau+1)}, z_{j+1}^{(\tau)}, \dots, z_M^{(\tau)})$ .
  - $\vdots$
  - Sample  $z_M^{(\tau+1)} \sim p(z_M | z_1^{(\tau+1)}, z_2^{(\tau+1)}, \dots, z_{M-1}^{(\tau+1)})$ .

# Check of Gibbs Sampling

- Gibbs Sampling が欲しい分布のサンプリングを与えることを簡単にチェック
- $p(\mathbf{z})$  が不変分布であること (← 詳細釣り合い)
  - サンプリングで  $\mathbf{z}_{\setminus i}$  の値は変更されないので周辺分布  $p(\mathbf{z}_{\setminus i})$  は不変
  - 定義より各ステップは  $p(z_i | \mathbf{z}_{\setminus i})$  からサンプリングを行うので、合わせて  $p(\mathbf{z})$  は不変である
- エルゴード性
  - エルゴード性の十分条件は条件付き分布の確率が 0 となる場合がないこと (これが成り立てば  $\mathbf{z}$  空間上の任意の点から別の任意の点へ有限回のステップで移動できる)
  - これが成り立たない場合、エルゴード性は明示的に証明される必要がある

# Relation to the Metropolis-Hastings Algorithm

- Metropolis-Hastings algorithm で、 $\mathbf{z}$  から  $\mathbf{z}^*$  への遷移を以下のようにする

$$q_k(\mathbf{z}^* | \mathbf{z}) = p(z_k^* | \mathbf{z}_{\setminus k}).$$

- このとき、 $p(\mathbf{z}) = p(z_k | \mathbf{z}_{\setminus k})p(\mathbf{z}_{\setminus k})$  と  $\mathbf{z}_{\setminus k}^* = \mathbf{z}_{\setminus k}$  より

$$\begin{aligned} A(\mathbf{z}^*, \mathbf{z}) &= \frac{p(\mathbf{z}^*)q_k(\mathbf{z} | \mathbf{z}^*)}{p(\mathbf{z})q_k(\mathbf{z}^* | \mathbf{z})} \\ &= \frac{p(z_k^* | \mathbf{z}_{\setminus k}^*)p(\mathbf{z}_{\setminus k}^*)p(z_k | \mathbf{z}_{\setminus k}^*)}{p(z_k | \mathbf{z}_{\setminus k})p(\mathbf{z}_{\setminus k})p(z_k^* | \mathbf{z}_{\setminus k})} = 1. \end{aligned}$$

となり、Metropolis-Hastings algorithm はサンプルを常に受理することがわかる

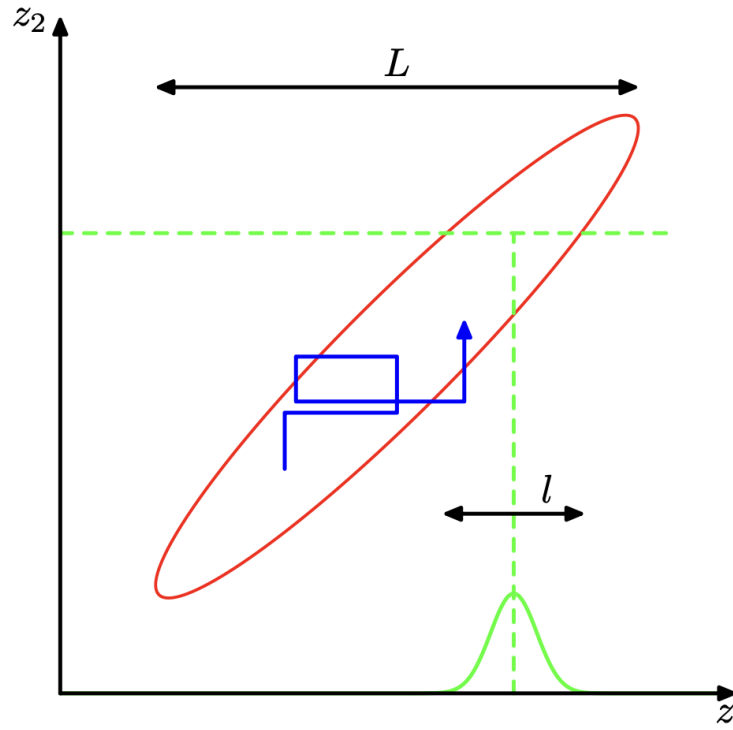


# Illustration of Gibbs Sampling

- 条件付き分布の幅を  $l$  とすれば、独立なサンプルを得るためのステップはおよそ  $(L/l)^2$  になる

**Figure 11.11**

Illustration of Gibbs sampling by alternate updates of two variables whose distribution is a correlated Gaussian. The step size is governed by the standard deviation of the conditional distribution (green curve), and is  $O(l)$ , leading to slow progress in the direction of elongation of the joint distribution (red ellipse). The number of steps needed to obtain an independent sample from the distribution is  $O((L/l)^2)$ .



# Over-Relaxation

- Gibbs sampling におけるランダムウォーク的な振る舞いを提言するアプローチの一つが**過剰緩和** (over-relaxation) である
- 条件付き分布がガウス分布である場合を考えると、ギブスサンプリングの各ステップで要素  $z_i$  の条件付き分布はある平均  $\mu_i$  と分散  $\sigma_i^2$  を持つ
- それを平均 0、分散 1 のガウス分布に従う確率変数  $\nu$  と  $-1 < \alpha < 1$  を用いて

$$z'_i = \mu_i + \alpha(z_i - \mu_i) + \sigma_i(1 - \alpha^2)^{1/2}\nu,$$

と置き換えることで、ステップの平均方向にバイアスをかける

- こうすることで、(変数が高い相関を持つとき) 状態空間内の方向付けられた移動を促進することができる

# Gibbs Sampling の実用性

- ギブスサンプリングが適応可能か否かは条件付き分布からのサンプリングが容易にできるかどうかによる
- ギブスサンプリングにあたって考慮すべき変数はマルコフブランケットで表される
  - マルコフブランケットで条件付けられた分布は対数凹関数になりがちなため、適応棄却サンプリングできる
  - グラフが指数分布族で表され共役性があるなら関数形が容易にわかるので標準的なサンプリング手法を使える
- 数列の相関を下げるため、1つの変数だけでなく複数の変数をまとめて更新する  
**ブロック化ギブスサンプリング**アルゴリズムというものもある

## 11.4 Slice Sampling

- 分布の特徴にあわせた適応的なステップサイズの選び方
- 1 変数の場合、 $z$  を付加的な変数  $u$  で拡張し、以下の確率分布

$$\hat{p}(z, u) = \begin{cases} 1/Z_p & (0 \leq u \leq \tilde{p}(z)) \\ 0 & (\text{otherwise}) \end{cases}$$

から一様にサンプリングして  $u$  を無視することで  $p(z)$  からサンプリングできる

- 具体的には、適当な  $z$  を初期値として以下を繰り返す
  1.  $\tilde{p}(z)$  を求め一様に  $u$  をサンプリングする
  2.  $u$  を固定し、 $\{z : \tilde{p}(z) > u\}$  から  $z$  を一様にサンプリングする

# Illustration of Slice Sampling

- 実際のように分布のスライスから直接サンプリングすることが難しい場合、代わりに  $\hat{p}(z, u)$  を不変にするサンプリングスキームを定義し、 $z^{(\tau)}$  を含む適当な区間から  $z^{(\tau+1)}$  を決定する

