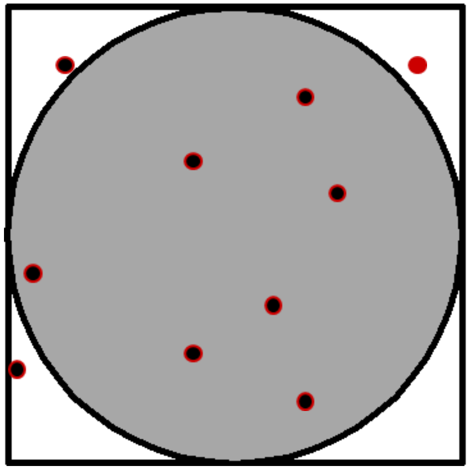
MCMC チュートリアル
入門から多峰性分布の扱いとその応用まで

Q

「マルコフ連鎖モンテカルロ法」って統計物理とかで昔から使ってるモンテカルロ法とは違う新しい方法なんですね？

A

いいえ同じです．名前が変わっただけ．試行が独立なタイプの「モンテカルロ法」とは違います．



マルコフ連鎖モンテカルロ法とは呼ばれないモンテカルロ法の例
（各試行が独立）

円の中に入った個数÷全体の個数

マルコフ連鎖モンテカルロ法（MCMC)

１９５０年代前半に誕生
電子計算機の誕生　⇒
　多数の原子や分子の振る舞いの計算

「（動的）モンテカルロ法」

MCMCという名前は他の分野（統計科学など）
で使われるようになってからのもの

MCMCの応用の展開

- @ 1950s~
  液体⇒磁性⇒素粒子（経路積分）
- @ 1990s~ (散発的に1980s)
  ベイズ統計を中心とする
  統計学・情報処理の諸分野 1990s~
- @ Next？
  最後に「他の応用」の話をします

Q

MCMCにデータを突っ込むと解析してくれる？

A

だめです．「モデル」（ベイズ）を書く必要がある

MCMCは計算の道具にすぎない

便利なツールはできています

(モデルを入れるとMCMCを実行)

Stan JAGS WinBUGS ..

統計データ解析とMCMC

MCMCだけ勉強するのは水道の蛇口だけ付けて配管をしないのと同じ

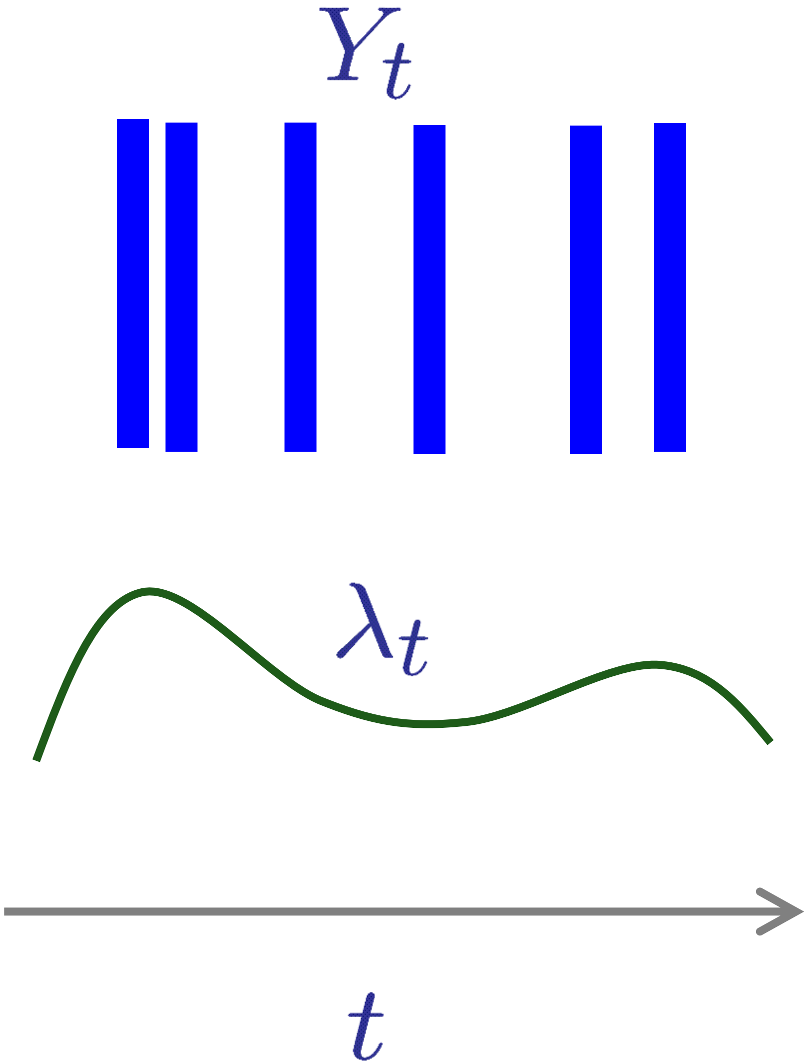
データ

$$\boxed{Y_t} \sim Poisson(\lambda_t Z)$$

$$\log \lambda_t = \beta_t$$

$$\beta_t = \beta_{t-1} + \epsilon_t$$

ノイズ



```
data {
  int T;
  int Y[T];
  int Z;}

parameters {
  real beta[T];
  real<lower=0>
s_mu;}

transformed parameters {
  real prob[T];
  for(t in 1:T)
    prob[t,1]<-exp(beta[t]); }

model {
  for(t in 3:T)
    beta[t] ~ normal(beta[t-1], s_mu);
  for(t in 1:T)
    Y[t]~poisson(Z*¥prob[t]);  }
```

STANのコード

（R,Pythonから呼び出す）

Q

要するに最適化の方法でしょう？

A

違います

ＭＣＭＣの目的

高次元（多変量）の分布からのサンプリング法

（＝　乱数の生成法）

@ 最適化手法ではない

（最適化に使うことはできる： simulated annealing）

@「高次元積分法」でもない（工夫すれば使えるが）

@「普通」のモンテカルロと最適化の中間の性質

今日の話

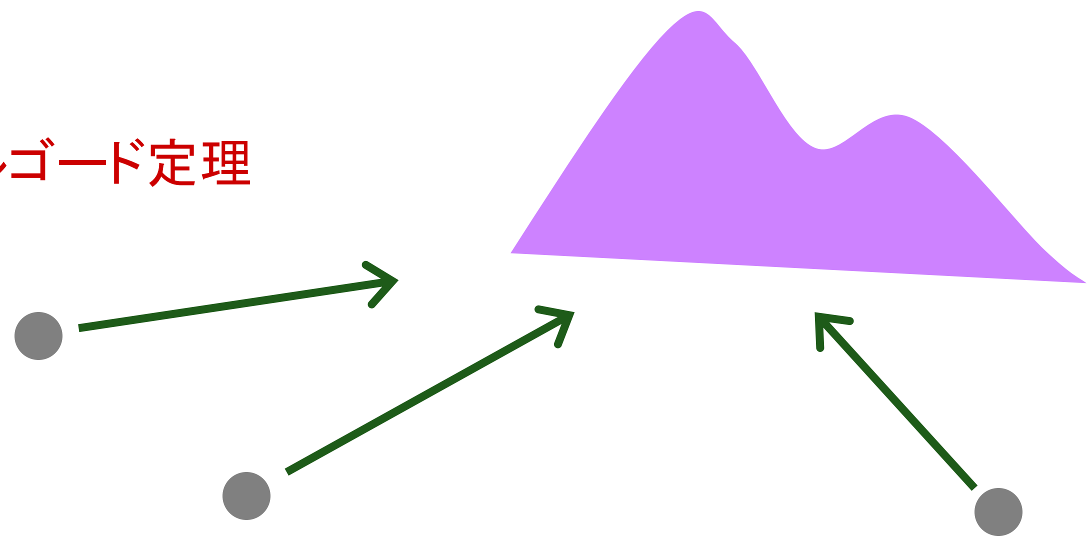
- MCMCの大雑把な話
  （原理からレプリカ交換法まで）
- ベイズ
  （ちょっとだけですからね！）
- レアイベントサンプリング

ＭＣＭＣの要点

サンプリングしたい分布に対して
それを「定常分布」とするマルコフ連鎖を構成

⇒　擬似乱数を使ってシミュレート

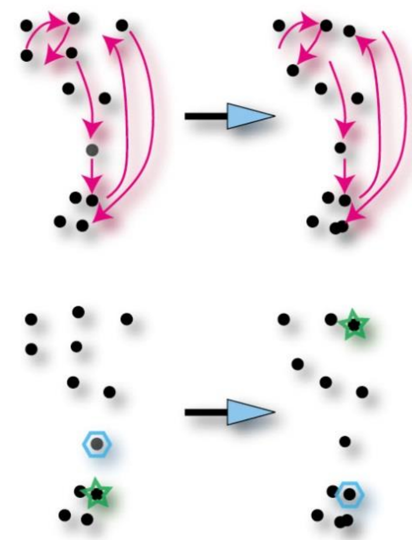
マルコフ連鎖のエルゴード定理



マルコフ連鎖の定常分布(不変分布)

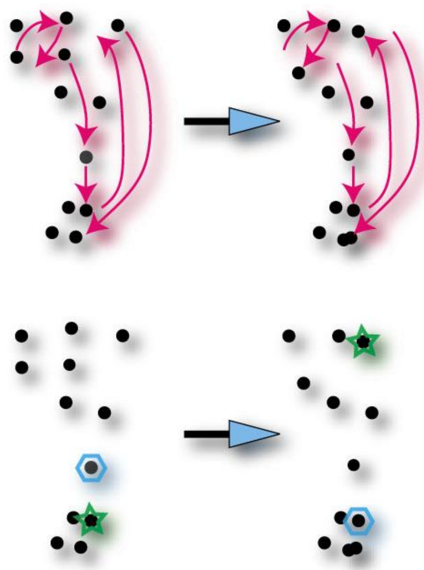
$P(x)$　多数の状態(状態空間の代表点)の分布

$$\sum_x P(x)\pi(x \to x') = P(x')$$



マルコフ連鎖の収束定理(エルゴード性)

「与えられた分布を定常分布に持つ」



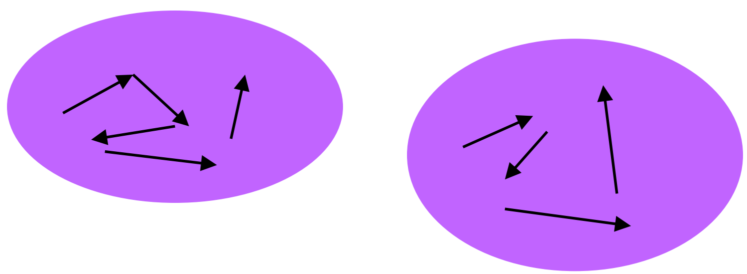
与えられた分布→与えられた分布　○

任意の分布→正しい分布　？

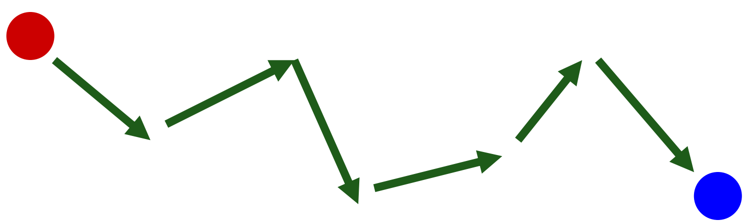
簡単な条件のもとで証明できる

条件

まず，複数の部分にわかれていたのではだめ

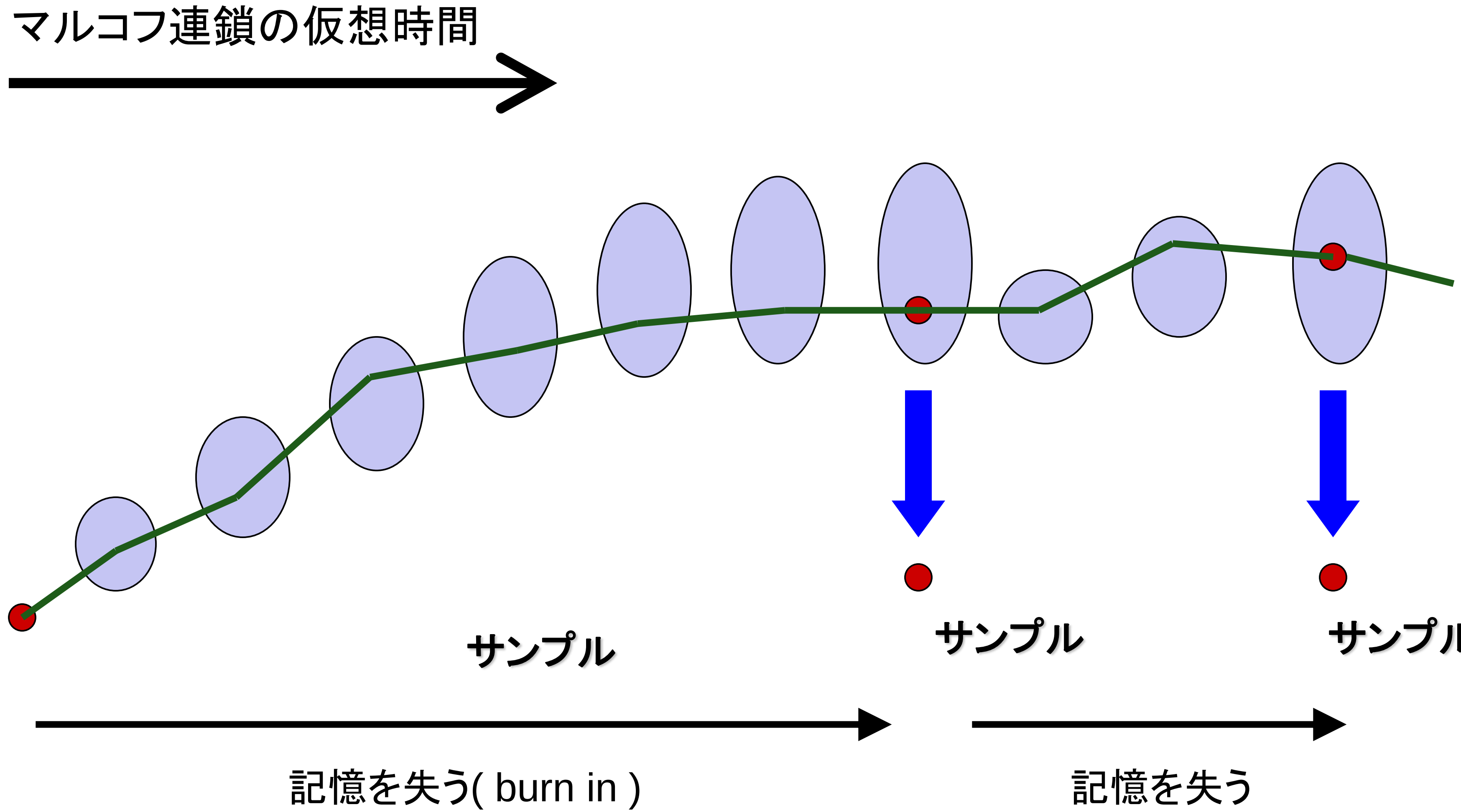


既約性

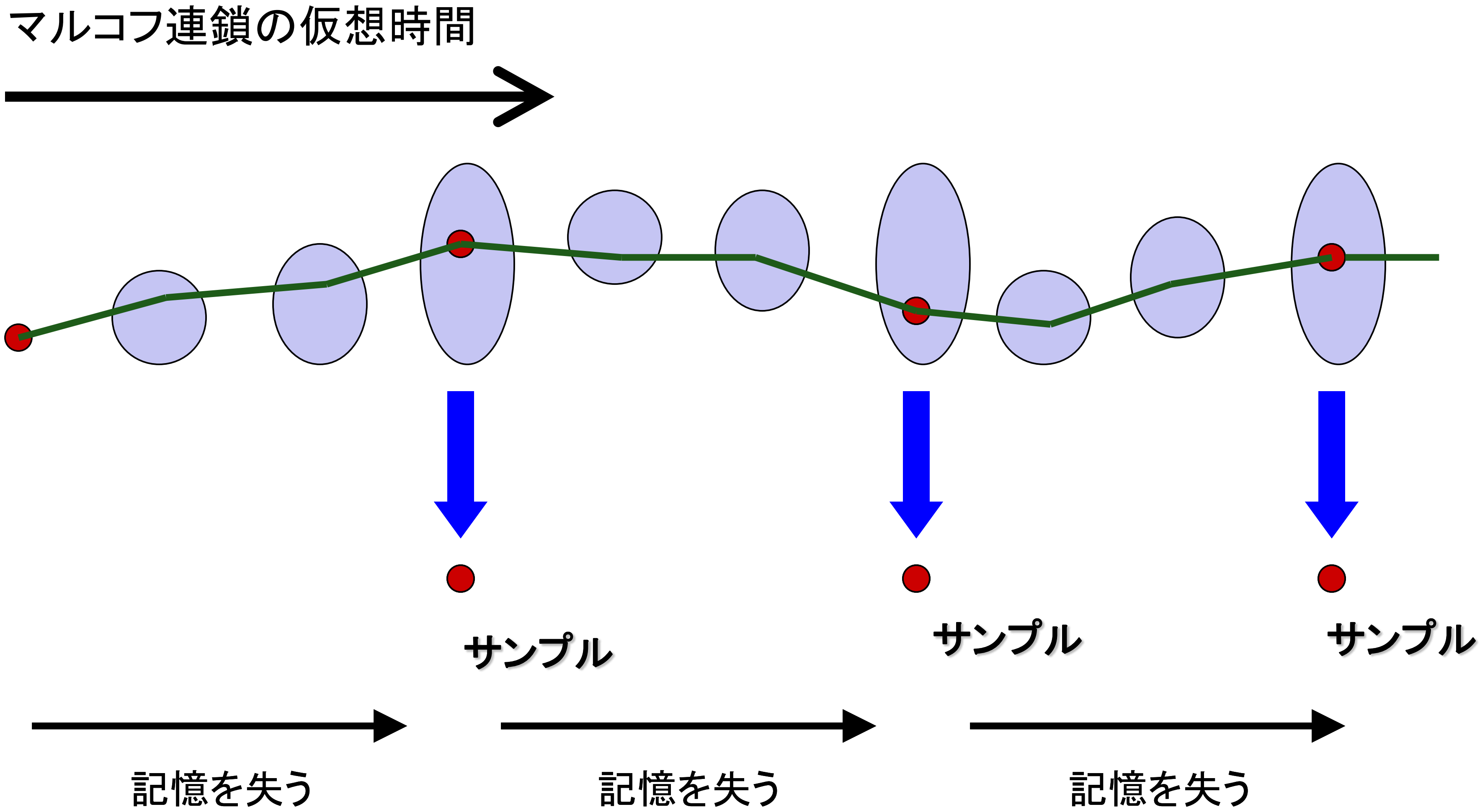


（＋α）

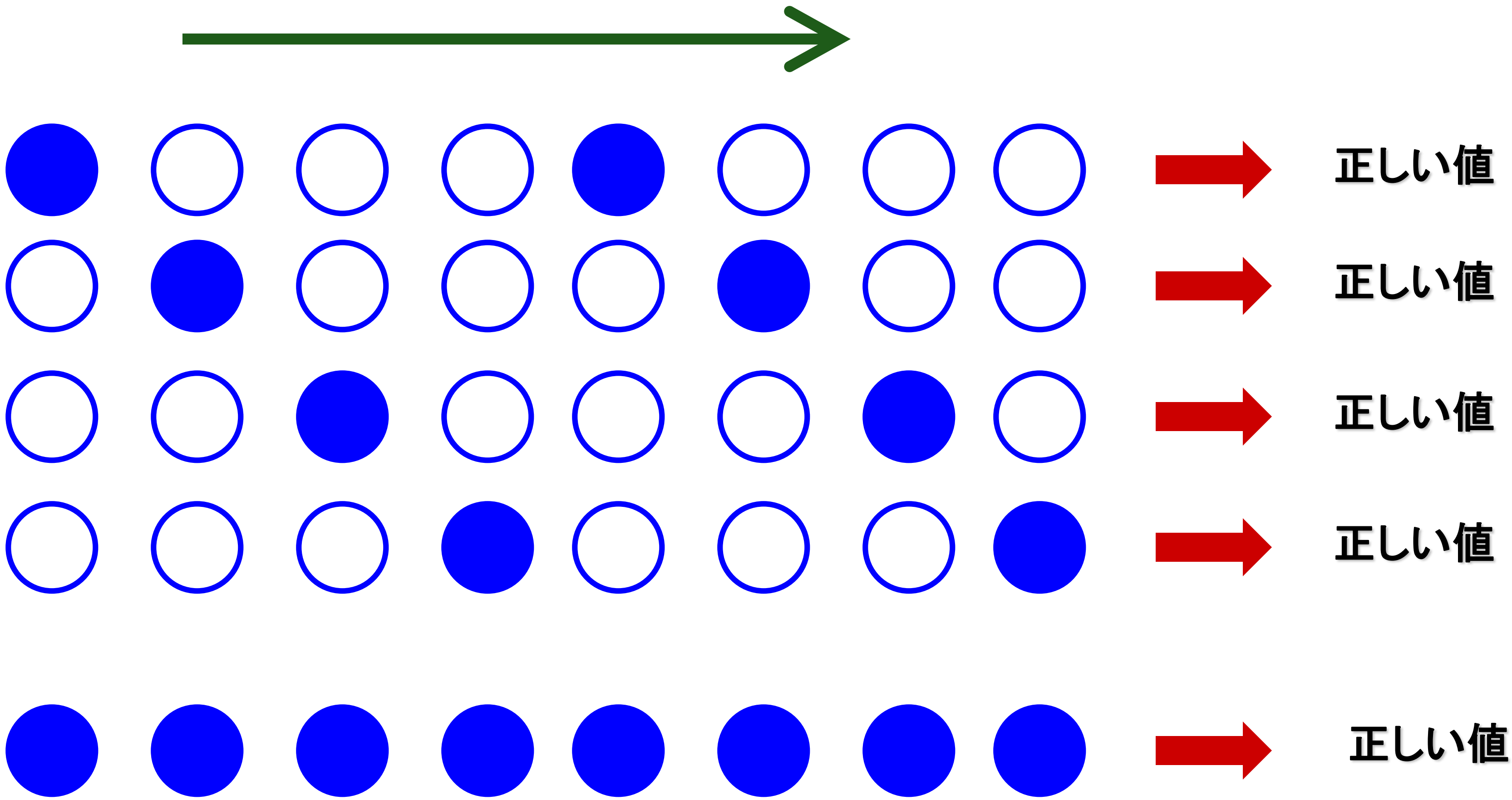
離れた初期状態から



多数のサンプルを採る



平均値の計算では飛ばす必要はない



分散などの不偏推定・・相関の補正が必要（高度）

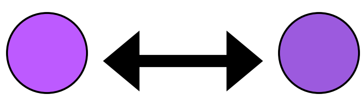
マルコフ連鎖を設計する

与えられた分布を定常分布にする
マルコフ連鎖は無数に存在する

設計の原理のひとつ　⇒　詳細釣り合い
(detailed balance)

詳細釣り合い（可逆性）

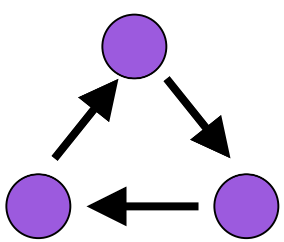
$$P(x)\pi(x \to x') = P(x')\pi(x' \to x)$$



各ペアについてつりあう

$$\sum_x P(x)\pi(x \to x') = P(x')$$

定常分布に
なる十分条件



詳細釣り合いを満たさない例

最近は「詳細釣り合いなしMCMC」もある

メトロポリス法

対称な提案分布 $Q(x, x') = Q(x', x)$

x の候補 x^{new}を

提案分布 $Q(x, x^{new})$ に従って生成

一様乱数 $0 \leq r < 1$ を発生

$r < \dfrac{P(x^{new})}{P(x)}$ なら x をx^{new} で置き換える

さもなければ何もしない

詳細釣り合いを満たすことの証明

２つの状態をα, γとする

$P(\alpha) \leq P(\gamma)$ として一般性を失わない

α→γ

$$r < 1 \leq \frac{P(\gamma)}{P(\alpha)} \qquad Q(\alpha,\gamma)$$

γ→α

$$r < \frac{P(\alpha)}{P(\gamma)} \leq 1 \qquad Q(\gamma,\alpha)\frac{P(\alpha)}{P(\gamma)}$$

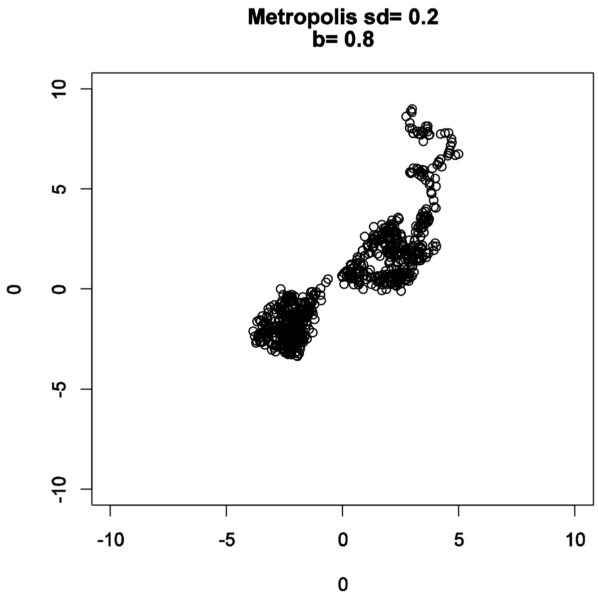
続き

$$P(\alpha)\pi(\alpha \to \gamma) = P(\gamma)\pi(\gamma \to \alpha)$$

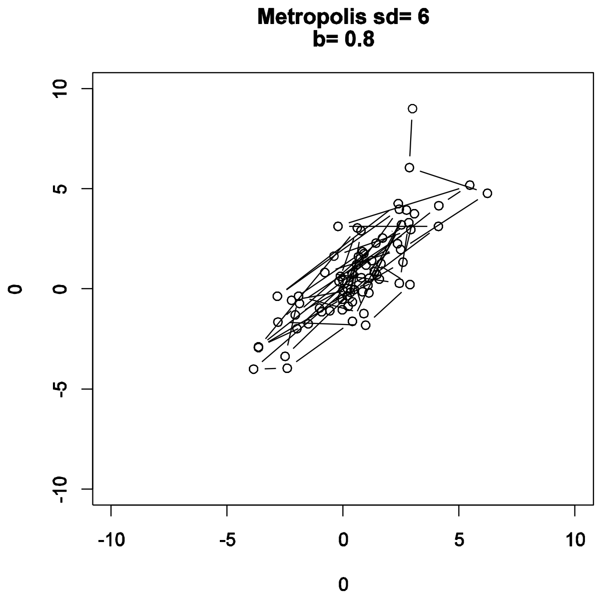
$$P(\alpha) \times Q(\alpha, \gamma) = P(\gamma) \times Q(\gamma, \alpha)\frac{P(\alpha)}{P(\gamma)}$$

$$Q(\alpha, \gamma) = Q(\gamma, \alpha)$$

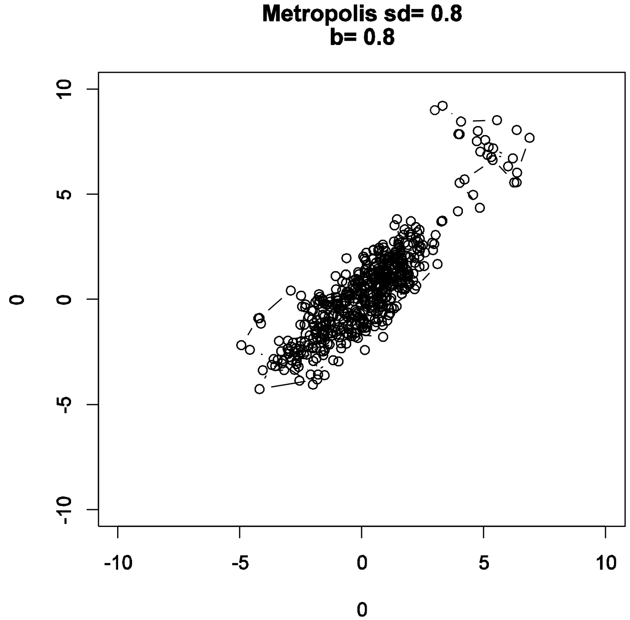
ステップ幅の影響



Metropolis sd= 0.2
b= 0.8
10
5
0
-5
-10
-10
-5
0
5
10
0
0



Metropolis sd= 6
b= 0.8
10
5
0
-5
-10
-10
-5
0
5
10
0
0



Metropolis sd= 0.8
b= 0.8
10
5
0
-5
-10
-10
-5
0
5
10
0
0

マルコフ連鎖の作り方

■メトロポリス法

■Gibbs Sampler（熱浴法）

➔メトロポリス・ヘイスティングス法（MH法)

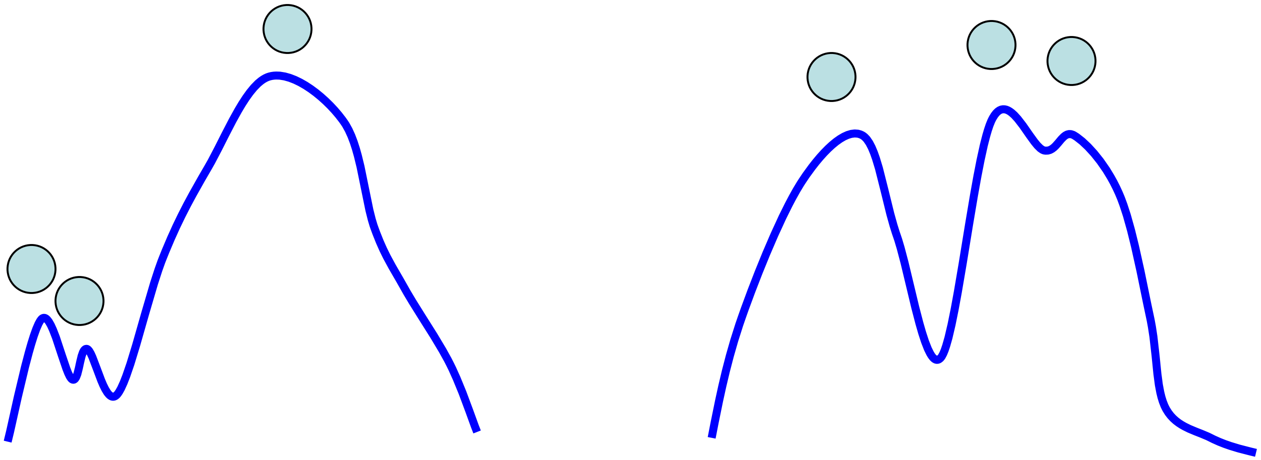
（上の２つの両方を含む）

■ハミルトニアンモンテカルロ法（Stanに採用)

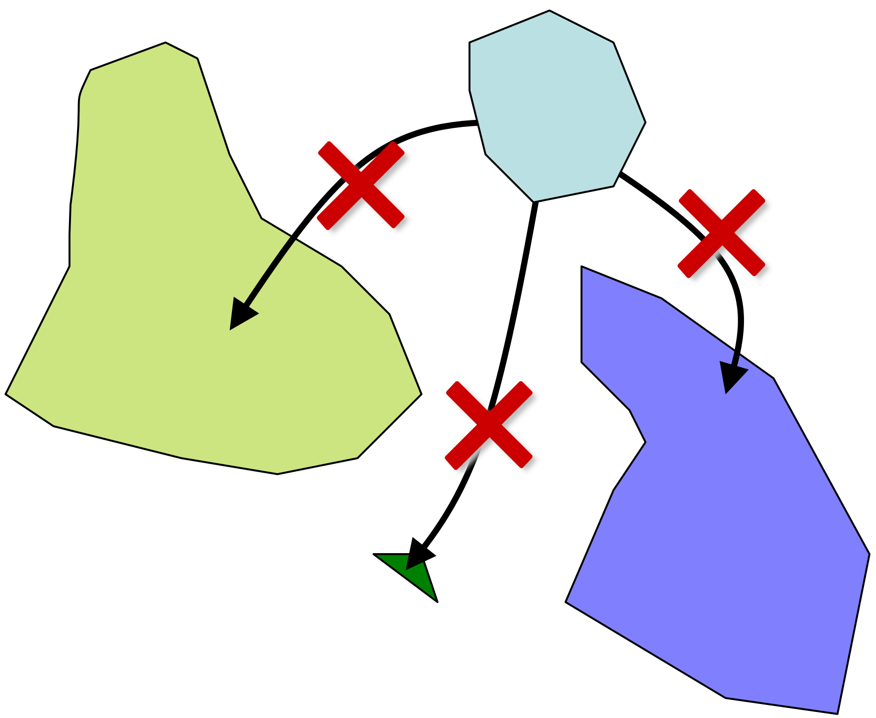
メトロポリス法の1stepをHamilton力学系で

■emcee

MCMCが困難に陥る場合



確率密度の大きいところが複数に分かれる



レプリカ交換モンテカルロ法
（パラレル・テンパリング）

たくさんの系を並列にシミュレートする

$$\{P_k(x^k)\}, k = 1, 2, \ldots, K$$

「たくさんの系」の作り方

それぞれ温度が少しづつ段階的に違うようにする

$$P_k(x^k) = \frac{\exp(-\beta_k E(x^k))}{Z(\beta_k)}$$

任意のパラメータ（β）についての族でもよい

$$P_k(x^k) = P(x^k|\beta_k)$$

交換操作

一定の規則で隣同士の状態を交換する

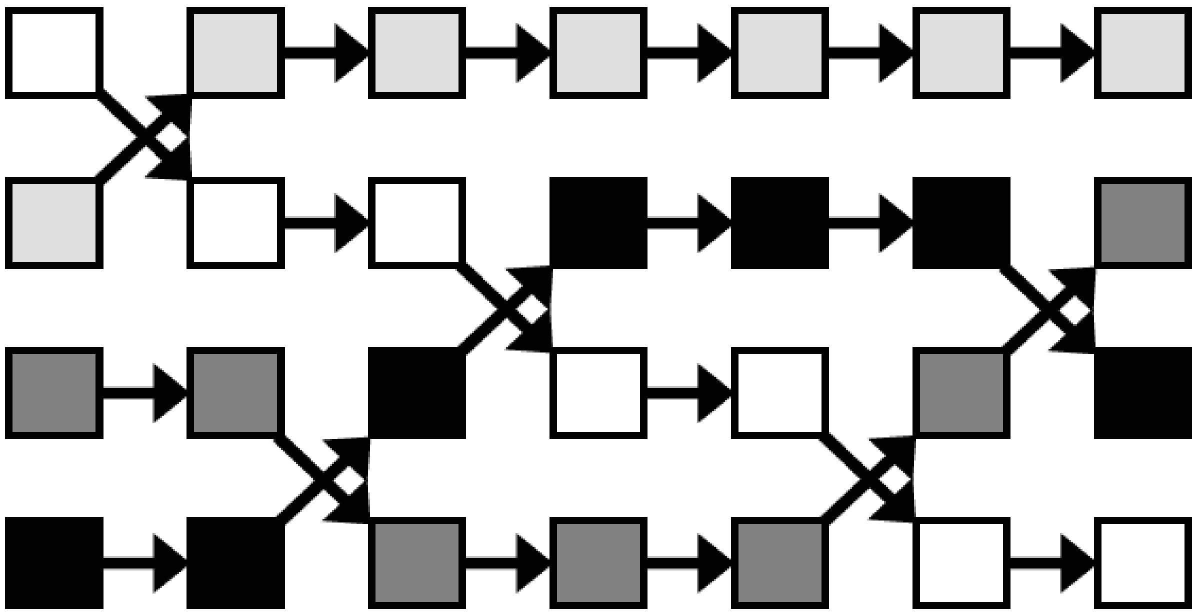
$$P_k(x^k) \qquad P_{k+1}(x^{k+1})$$

$$x^{k^{(t+1)}} \leftarrow x^{k+1^{(t)}}$$

$$x^{k+1^{(t+1)}} \leftarrow x^{k^{(t)}}$$

交換の様子

（系４つの場合）



交換の規則

- 以下の比を計算

$$W = \frac{P_{k+1}(x^{k^{(t)}})\,P_k(x^{k+1^{(t)}})}{P_{k+1}(x^{k+1^{(t)}})\,P_k(x^{k^{(t)}})}$$

- 一様乱数　0≦ *rnd* ＜1 を発生
  rnd ＜*W* なら交換する

同時分布についての詳細釣り合い

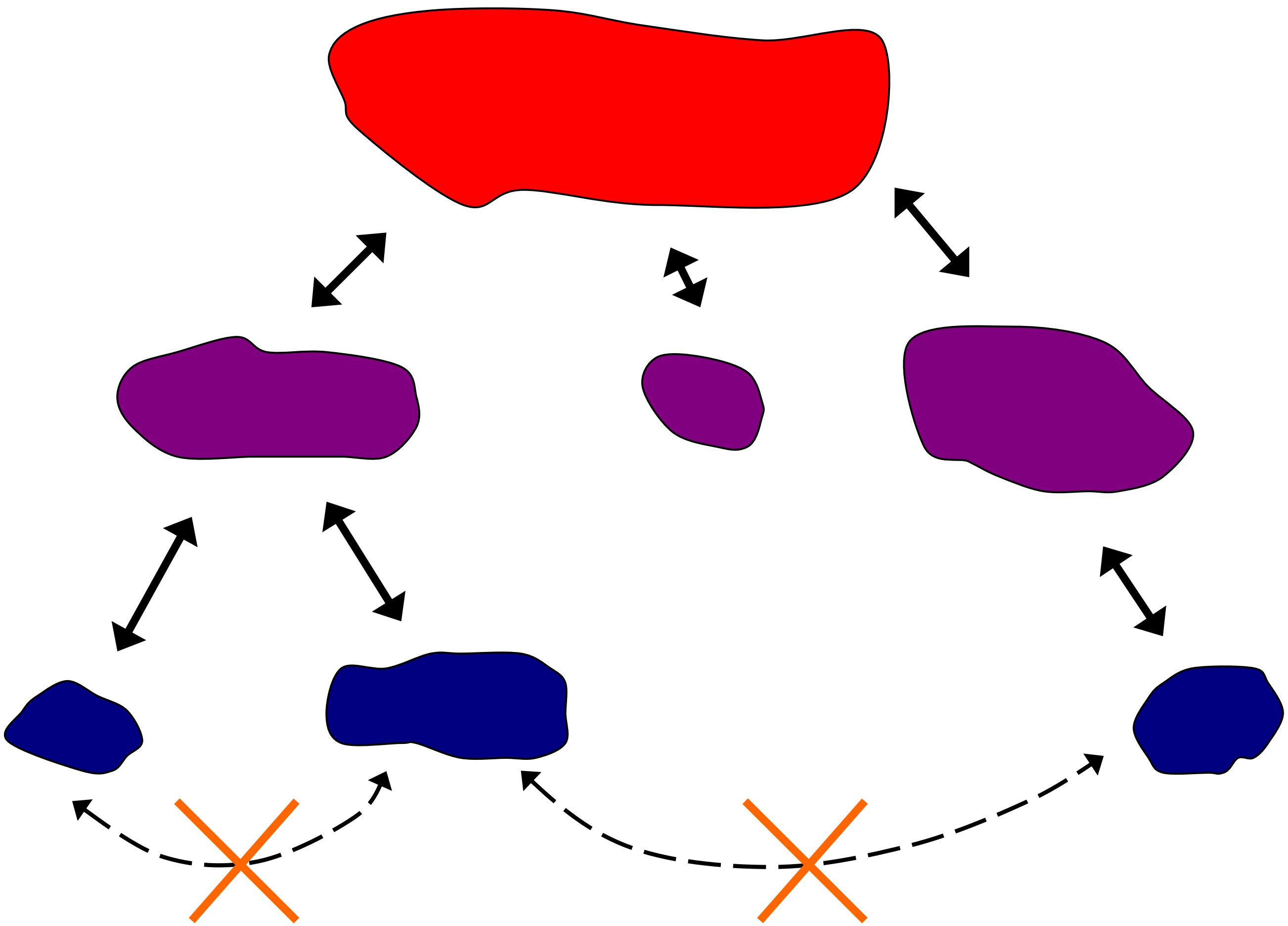
同時分布が交換操作を含めたマルコフ連鎖の
不変分布になっている

$$P(x^1, x^2, \ldots, x^K) = P_1(x^1)P_2(x^2)\ldots P_3(x^K)$$

「交換」の部分がメトロポリス法の試行に相当

$$W = \frac{P_1(x^1)\ldots \boxed{P_k(x^k)P_{k+1}(x^{k+1})} \ldots P_K(x^K)}{P_1(x^1)\ldots \boxed{P_{k+1}(x^{k+1})P_k(x^k)} \ldots P_K(x^K)}$$

交換の効果で橋をかける



ベイズ統計（ちょっとだけ）

岩波データサイエンス　全6巻　各1500円

1巻　ベイズ推論
2巻　自然言語処理
3巻　因果推論
4巻　地理空間情報
5巻　スパース推定
6巻　時系列解析(6月刊)


Wonderful R 2
StanとRで
ベイズ統計モデリング
松浦健太郎 著
共立出版


データ解析のための
統計モデリング入門
久保拓弥

時系列の状態空間モデル

データ発生のメカニズムをモデルとして仮定する

内部状態xの発展:システム方程式

$$x(t+1) = f\big(x(t)\big) + \eta(\mathrm{t})$$

雑音

観測してデータyを得る:観測方程式

$$y(t) = g\big(x(t)\big) + \epsilon(t)$$

雑音

xを生成,そこからyを生成 ➔ 逆にyを与えてxを推測

シミュレーション **データ解析**

内部状態xの発展

$$x(t+1) = f\big(x(t)\big) + \eta(\mathrm{t})$$

$$p(x) \quad x = (x(1), x(2), \dots)$$

観測してデータyを得る

$$y(t) = g\big(x(t)\big) + \epsilon(t)$$

$$p(y|x) \quad y = (y(1), y(2), \dots)$$

$$p(x, y) = p(y|x)p(x)$$

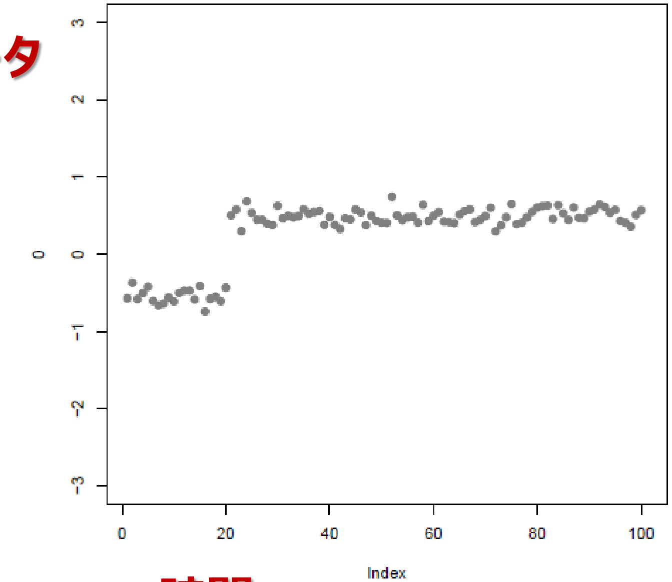
条件つき確率

$$p(x, y) = p(y|x)p(x)$$

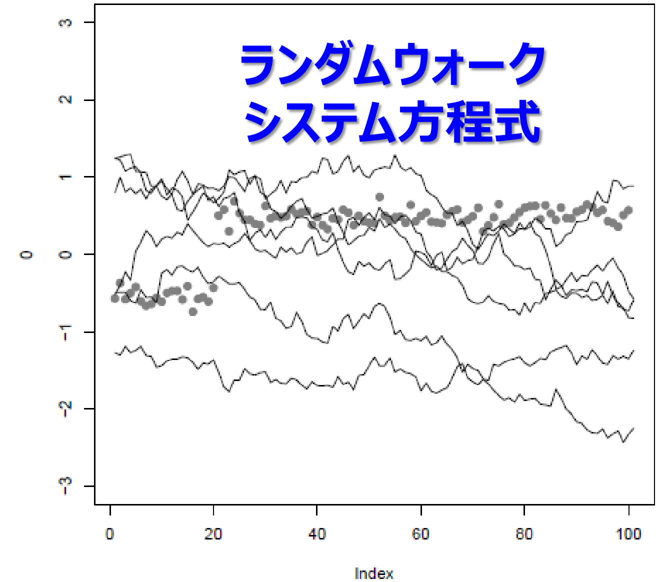
$$p(x|y) = \frac{p(x, y)}{p(y)} = \frac{p(y|x)p(x)}{\int p(y|x)p(x)dx}$$

ベイズの公式

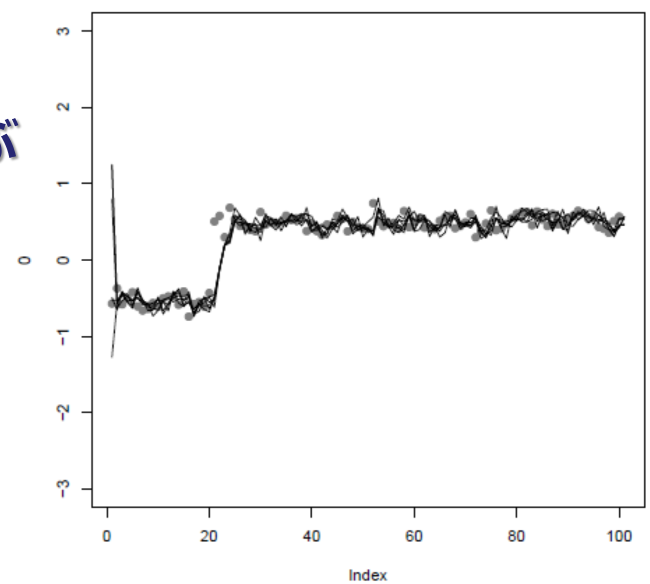
データ



時間→

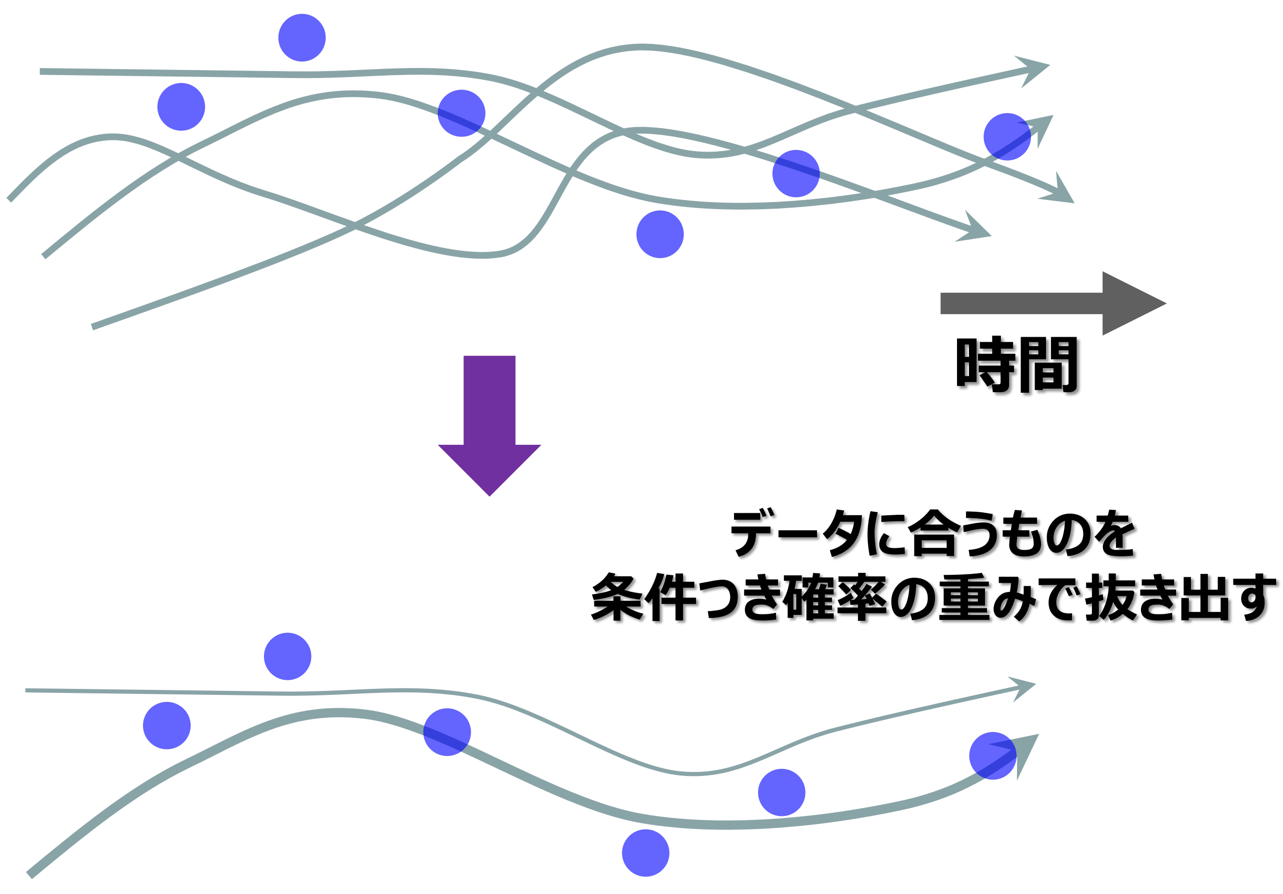


データに合う
ようなパスを選ぶ

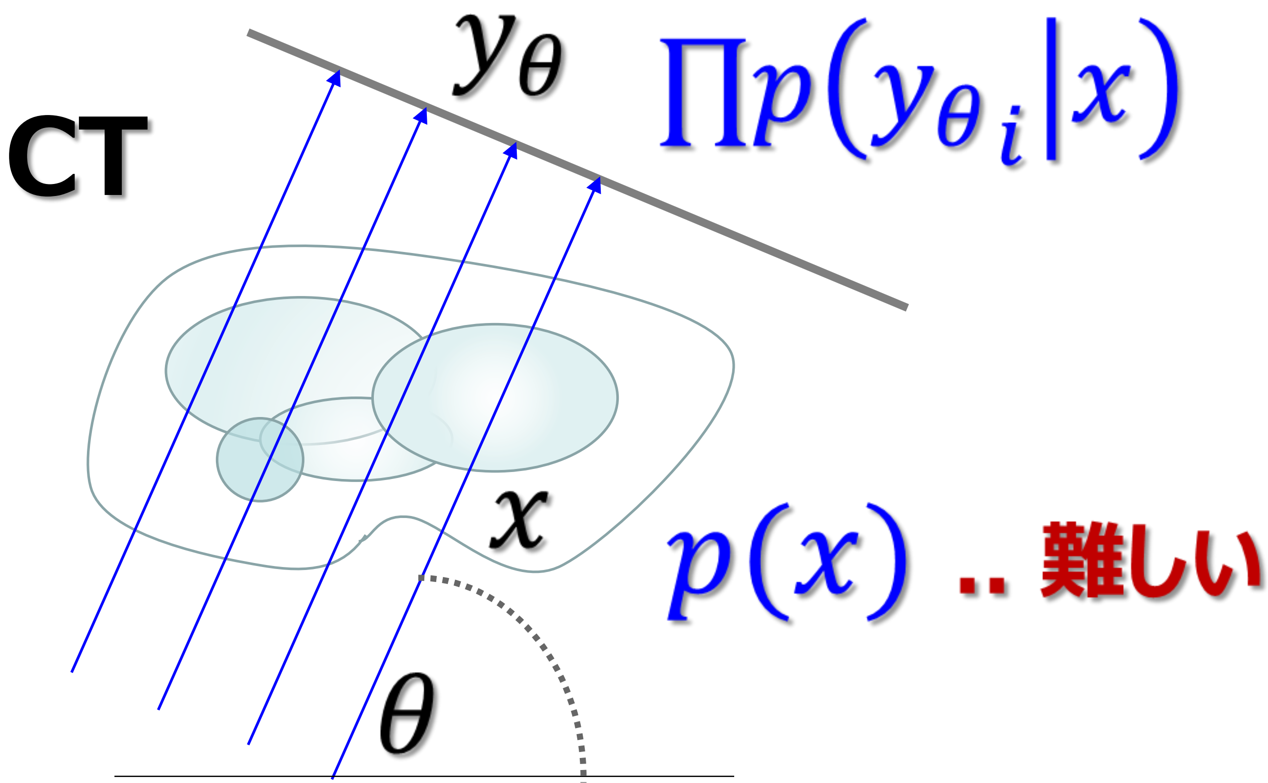


これがあてはめの
結果と考える
（誤差・不確実性を含む）

シミュレーションとデータ同化（一般化）



統計的逆問題も同様の枠組みで



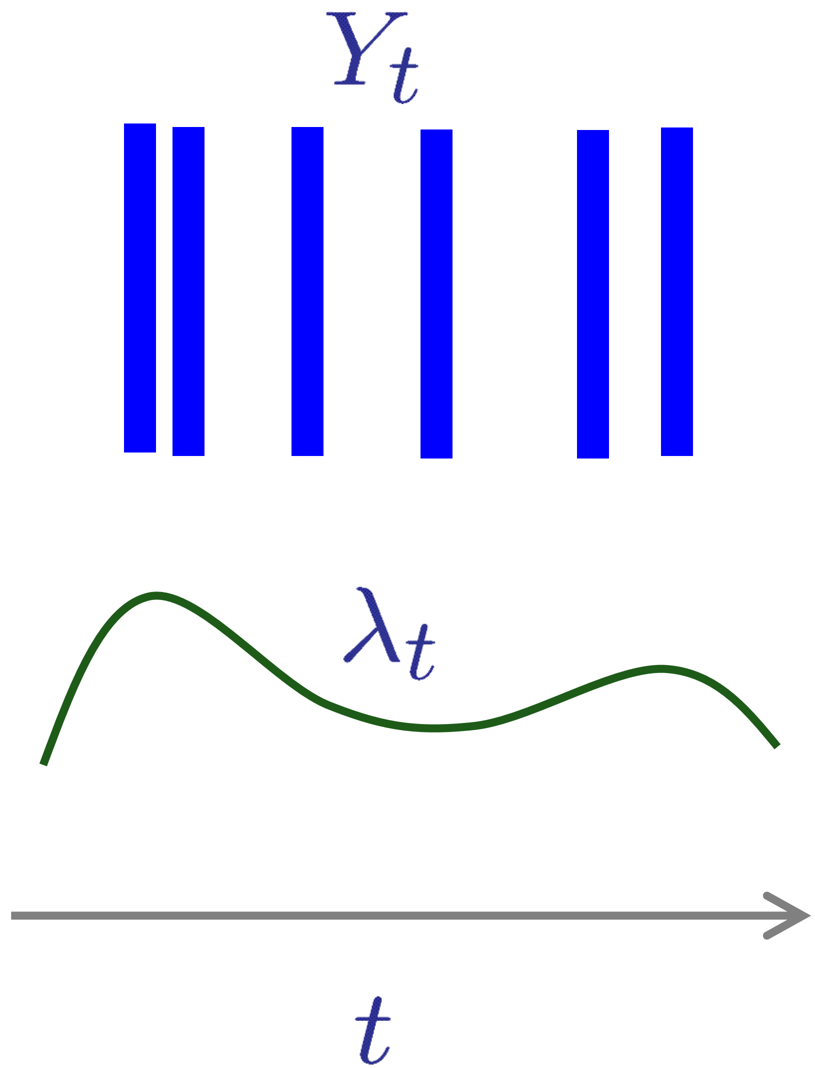
データ

$$Y_t \sim Poisson(\lambda_t Z)$$

$$\log \lambda_t = \beta_t$$

$$\beta_t = \beta_{t-1} + \epsilon_t$$

ノイズ



条件つき確率

$$p(x, y) = p(y|x)p(x)$$

$$p(x|y) = \frac{p(x, y)}{p(y)} = \frac{p(y|x)p(x)}{\int p(y|x)p(x)dx}$$

どうやって扱う？ ➔ MCMC!

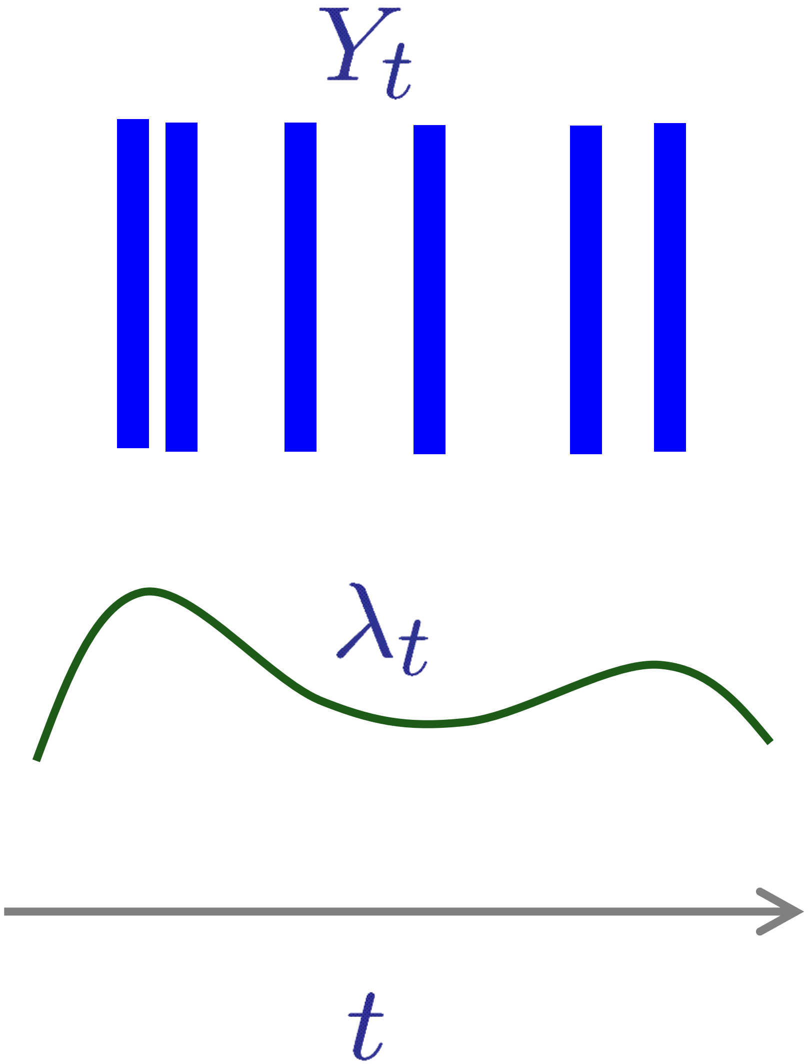
データ

$$Y_t \sim Poisson(\lambda_t Z)$$

$$\log \lambda_t = \beta_t$$

$$\beta_t = \beta_{t-1} + \epsilon_t$$

ノイズ



```
data {
  int T;
  int Y[T];
  int Z;}

parameters {
  real beta[T];
  real<lower=0>
s_mu;}

transformed parameters {
  real prob[T];
  for(t in 1:T)
    prob[t,1]<-exp(beta[t]); }
```

```
model {
  for(t in 3:T)
    beta[t] ~ normal(beta[t-1], s_mu);
  for(t in 1:T)
    Y[t]~poisson(Z*¥prob[t]);  }
```

STANのコード

（R,Pythonから呼び出す）

レアイベントサンプリング

Example 1 **Kitajima & Kikuchi (2015**

魔方陣の数を数える

RESEARCH ARTICLE

Numerous but Rare: An Exploration of Magic Squares

Akimasa Kitajima[1,2]*, Macoto Kikuchi[3,2]

Citation: Kitajima A, Kikuchi M (2015) Numerous but Rare: An Exploration of Magic Squares. PLoS ONE 10(5): e0125062. doi:10.1371/journal.pone.0125062

Pinn K, Wieczerkowski C (1998)
Paralell Tempering (Replica Exchange MCMC)

N＝30で6.56x10^2056 個

ハードな制約条件による困難

魔方陣をたくさん作りたいが・・

魔方陣から別の魔方陣を簡単に作る方法は思い浮かばない

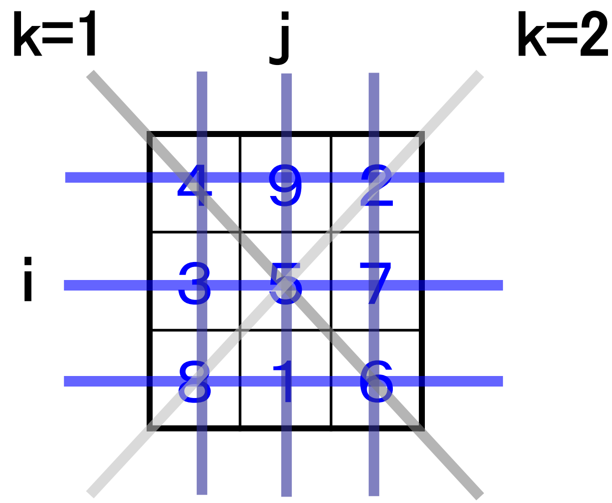
MCMCの言葉でいえば

「マルコフ基底」の選び方がわからない

（たぶん想像もつかないくらい複雑？）

制約条件をゆるめる

4	9	2
3	5	7
8	1	6



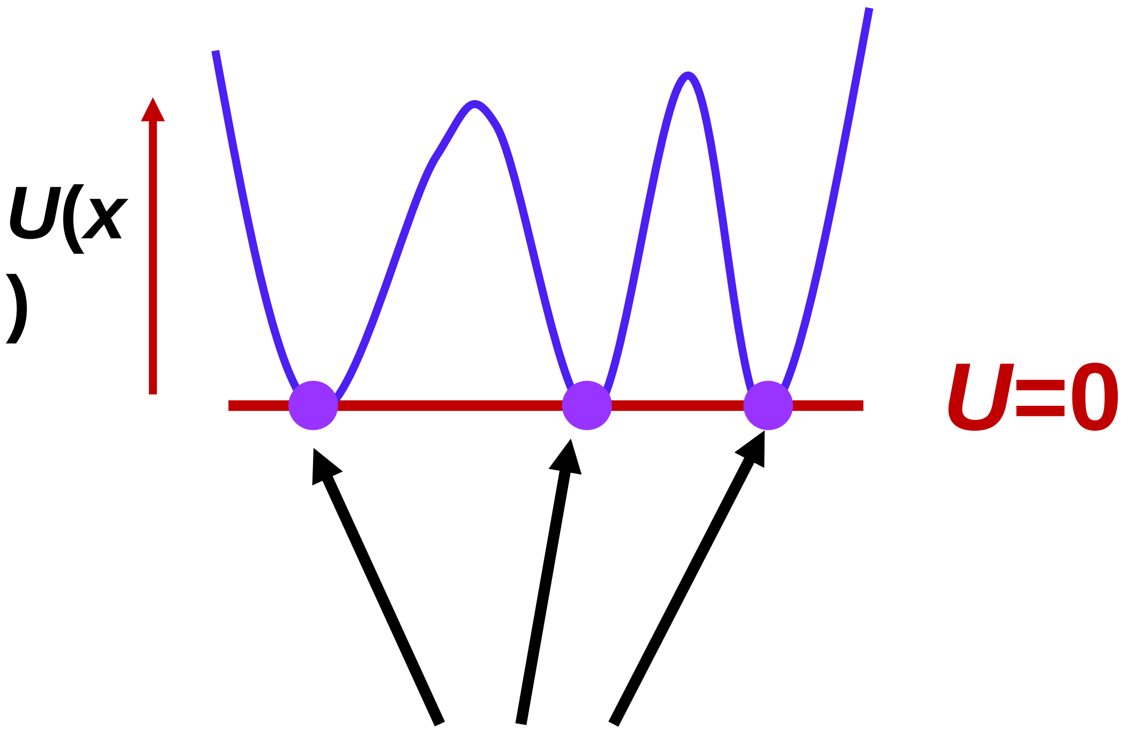
$$U(\{x_i\}) = \sum_i |rsum_i - C| + \sum_j |csum_j - C| + \sum_{k=1,2} |dsum_k - C|$$

rsum_i =行iの数字の和, *csum_j* =列jの数字の和,
dsum_k =対角線の数字の和 (対角線2本)

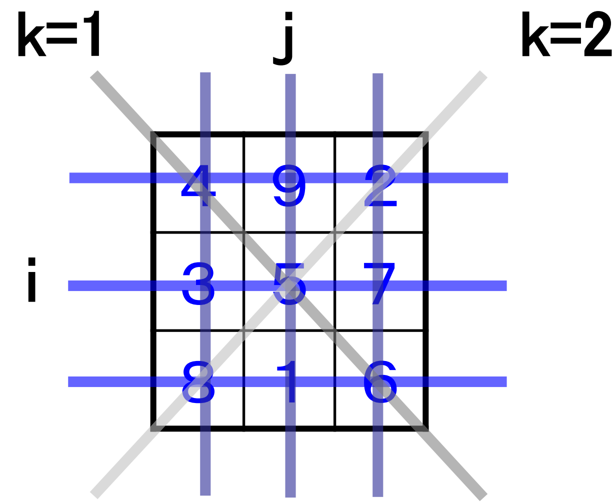
配置xが魔方陣になっている ⇔ $U(x)=0$

配置xが魔方陣になっている　⇔ $U(x)=0$

$$U(\{x_i\}) = \sum_i |rsum_i - C| + \sum_j |csum_j - C| + \sum_{k=1,2} |dsum_k - C|$$



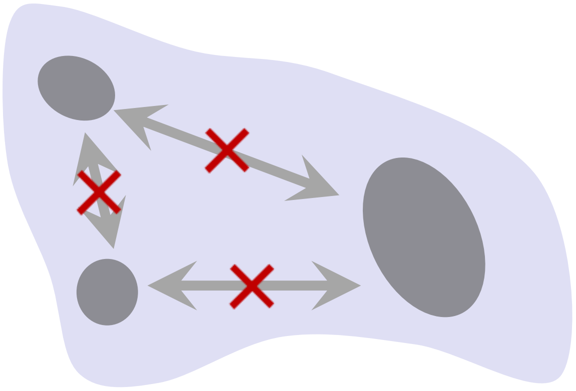
Complete Magic Square



MCMCを導入

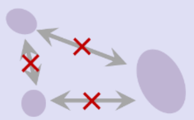
適当な1変数関数 $f(u) \geq 0$を定めて
重み$f(u)$でメトロポリス法のMCMCを走らす
（制約をゆるめてあるのでmoveは適当でよい）

一見，U=0のところにUの値が集中
するような重み$f(u)$を考えれば
ベストな気がする



しかしこれでは，うまくいかない
いろいろな魔方陣の間をつなぐことができない
（制約をゆるめないのと同じになってしまう）

重みの選び方が難しい



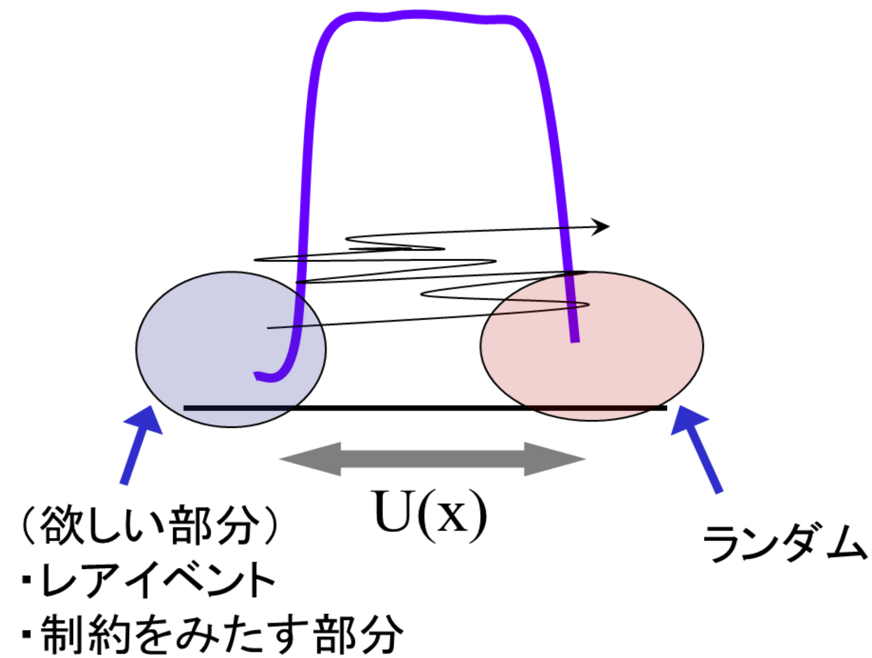
ではもっと広がった$f(u)$を考えれば
よいかというと，あっという間に分布が
広がってしまい，正しい魔方陣(U=0)
が出る数がゼロになってしまう
（次元の呪い；「軒を貸して母屋を取られる」）

マルチカノニカル法の $f(u)$ の選び方

マルチカノニカル法では
Uの値が大体一様に
分布するように$f(u)$を選ぶ
（5割やそこらの誤差は許容）

■「xがランダム」とは全く違う
（大数の法則！）

■「コイン100回振って
表100回と裏表半々が
等確率」みたいな感じにしたい



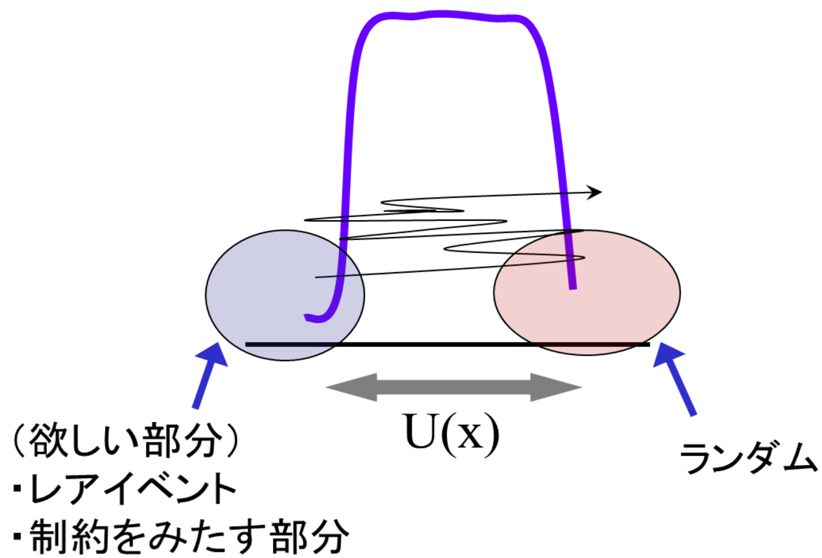
マルチカノニカル法の $f(u)$ の選び方（続き）

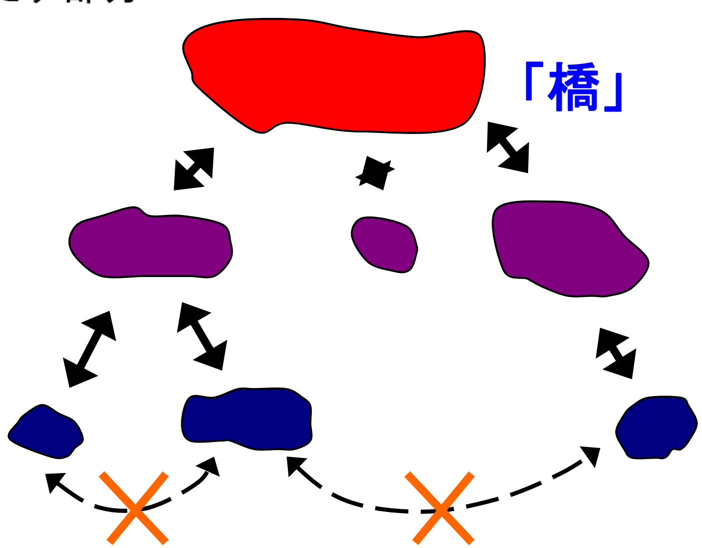
- この選び方ができたとして，なぜうまくいくか
- どうやって $f(u)$ を見つけるか

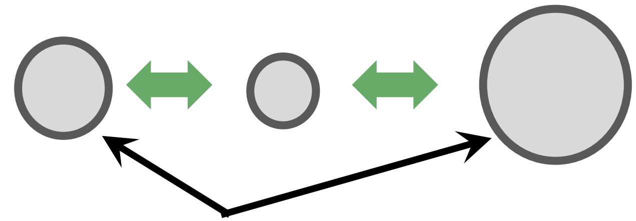
マルチカノニカル法の $f(u)$ の選び方（続き）

- この選び方ができたとして，なぜうまくいくか
- どうやって $f(u)$ を見つけるか

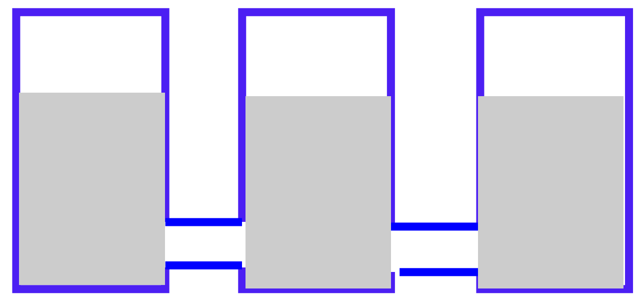
Uの値が広い範囲を動く利点（1）





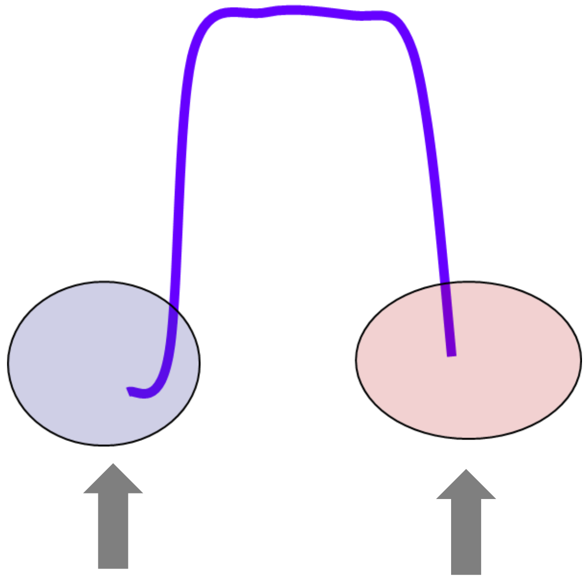


直接繋がっていなくても正しい比率



水槽と同じ理屈

Uの値が広い範囲を動く利点（2）



$f(0)$　$f(\text{ランダム})$

（欲しい部分）
・レアイベント
・制約をみたす部分

個数の見積もりがすぐにできる

$$\frac{f(0)}{f(\text{ランダム})} \times \text{総数}$$

× 完全に平らのとき

$$\frac{\text{実際の計算でU=0となった数}}{\text{完全に平らなときU=0となる数}}$$

完全に平らでない分の補正

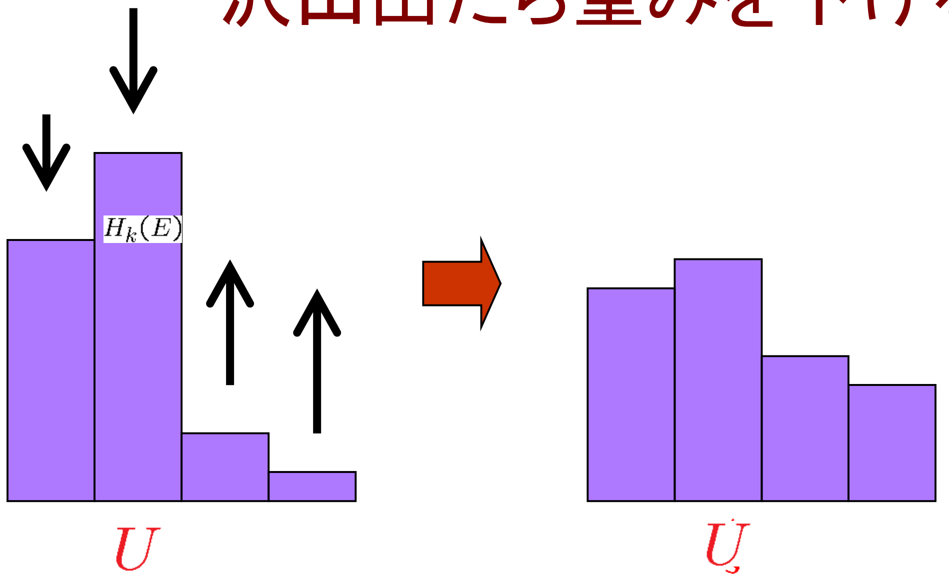
マルチカノニカル法の
$f(u)$ の選び方（続き）

- この選び方ができたとして，なぜうまくいくか
- どうやって$f(u)$ を見つけるか

最も簡単な学習法（Entropic Sampling）

[シミュレーション→ヒストグラム]　を繰り返す

沢山出たら重みを下げる

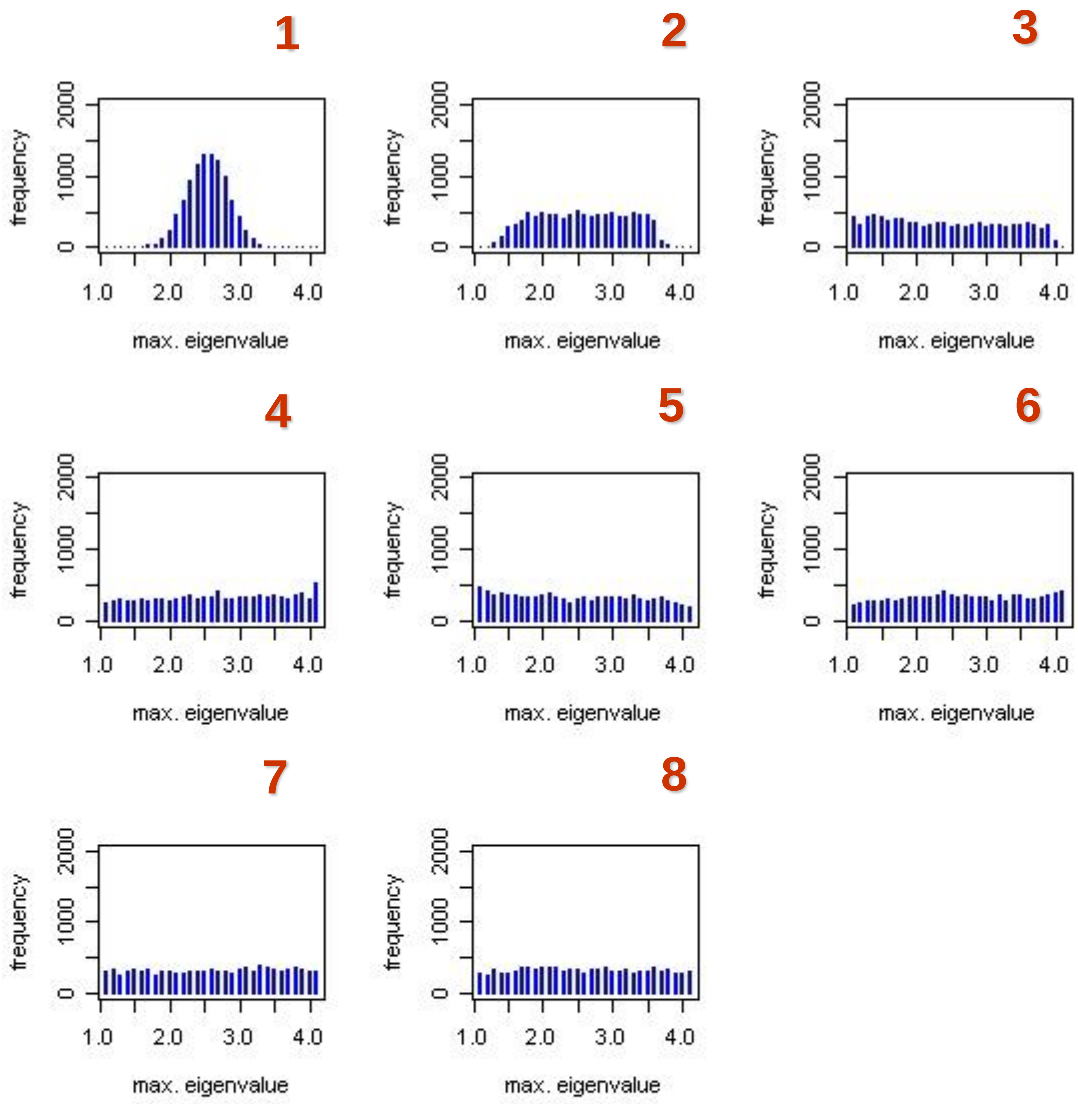


Histogram　$H_k(U)$

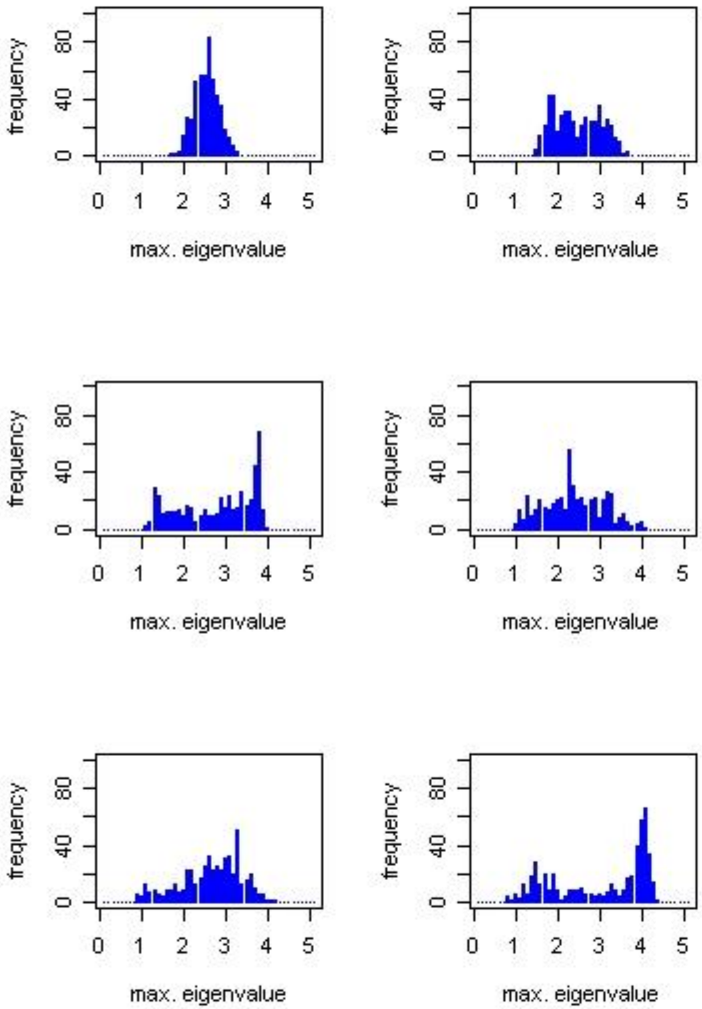
$$M_k = \frac{\sum_{\xi} H_k(U)}{\text{(number of bins)}}$$

$$f_{k+1}(U) = f_k(U) \div \frac{H_k(U)+\epsilon}{M_k}$$

ϵ　a constant, say 1.0


1
2
3
4
5
6
7
8
frequency
0
1000
2000
1.0
2.0
3.0
4.0
max. eigenvalue

An Example of Bad Convergence



(1) Request of a broad "flat" region
(2) Insufficent Sampling in each iteration
→ instability

Wang-Landau algorithm

1. Initialize Weights $w(E)$ Set $C \leftarrow 1/e$
2. Set/Reset Histogram $\boxed{H(E) \leftarrow 0}$
3. Metropolis-Hastings update with the weight $w(E(x))$
4. If the current state is x and $E = E(x)$

discount:	$w(E) \leftarrow w(E) * C$
increment:	$H(E) \leftarrow H(E) + 1$

5. If $H(E)$ becomes "sufficiently flat"

   $\boxed{C \leftarrow \sqrt{C}}$ and Goto 2 else Goto 3.
6. If C become sufficiently small then quit.

Example 2 $\xi =$ "chaoticity"

Chaotic Dynamical Systems

Search for rare initial conditions that gives rare trajectories

e.g., Coupled Standard Map

$$
\begin{aligned}
u_{n+1} &= u_n - \frac{K}{2\pi}\sin(2\pi v_n) + \frac{k}{2\pi}\sin(2\pi(v_n + y_n)) \\
v_{n+1} &= v_n + u_{n+1} \\
x_{n+1} &= x_n - \frac{K}{2\pi}\sin(2\pi y_n) + \frac{k}{2\pi}\sin(2\pi(v_n + y_n)) \\
y_{n+1} &= y_n + x_{n+1}
\end{aligned}
$$

Probability of regular trajectory (fragments) embedded in chaotic sea

Our method

- Initial Condition Sampling (deterministic systems)

  initial condition x_0

  Define "Mixing Time" $T_\epsilon(x_0)$

  Calculate the distribution $P(T_\epsilon)$

- Muticanonical (or Parallel Tempering)

  ➔ efficient sampling with correct prob.

Multicanonical Sampling

Simulation of the dynamical system

$$
\begin{aligned}
u_{n+1} &= u_n - \frac{K}{2\pi}\sin(2\pi v_n) + \frac{k}{2\pi}\sin(2\pi(v_n + y_n))\\
v_{n+1} &= v_n + u_{n+1}\\
x_{n+1} &= x_n - \frac{K}{2\pi}\sin(2\pi y_n) + \frac{k}{2\pi}\sin(2\pi(v_n + y_n))\\
y_{n+1} &= y_n + x_{n+1}
\end{aligned}
$$

Metropolis-Hastings iterations

Optimization of Weight by Wang-Landau (or entropic sampling) iterations

Results(Figures)

See **arXiv:1003.2013**

Multicanonical Sampling of Rare Trajectories in Chaotic Dynamical Systems

Akimasa Kitajima, Yukito Iba

Computer Physics Communications
Volume 182, Issue 1, Pages 1-280 (January 2011)