

2019年5月23日

分子モデリングと分子シミュレーション

分子動力学法とモンテカルロ法

東京大学大学院農学生命科学研究科

アグリバイオインフォマティクス

教育研究プログラム

寺田 透

本日の講義内容

- 運動方程式
- 分子動力学法
 - 課題1
- 実験データとの比較
- モンテカルロ法
 - 課題2
- 定温分子動力学法

古典力学の復習

- ニュートンの運動方程式

$$\mathbf{F} = m\mathbf{a} \quad \mathbf{F}: \text{力、} m: \text{質量、} \mathbf{a}: \text{加速度}$$

\mathbf{F} 、 \mathbf{a} は3次元ベクトル→太字で表記

m はスカラー→斜体で表記

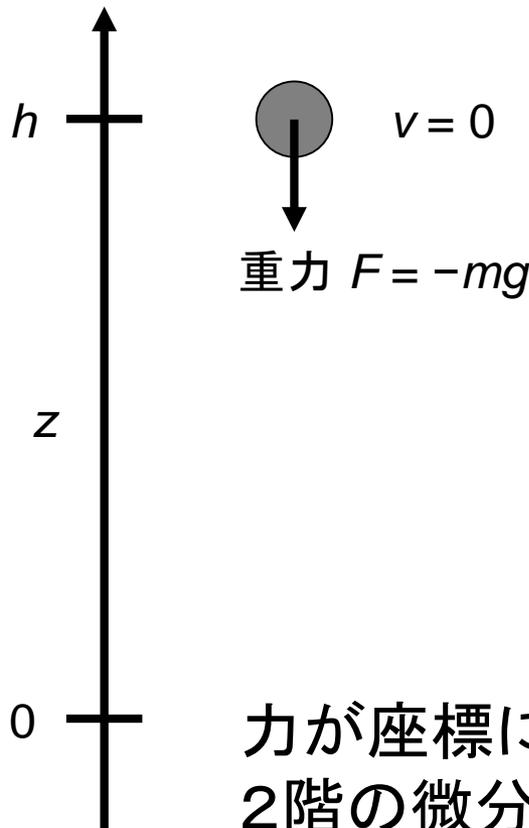
- 加速度 \mathbf{a} は座標 \mathbf{r} の時間 t に関する2次微分

$$\mathbf{v} = \frac{d\mathbf{r}}{dt} = \left(\frac{dx}{dt}, \frac{dy}{dt}, \frac{dz}{dt} \right)$$

$$\mathbf{a} = \frac{d\mathbf{v}}{dt} = \frac{d^2\mathbf{r}}{dt^2} = \left(\frac{d^2x}{dt^2}, \frac{d^2y}{dt^2}, \frac{d^2z}{dt^2} \right)$$

運動方程式の解(1)

- 質量 m の玉を高さ h から落とす



$$m \frac{d^2 z}{dt^2} = -mg \text{ より、} \frac{d^2 z}{dt^2} = \frac{dv_z}{dt} = -g,$$

$$v_z(t) = -gt + C, \quad v_z(0) = 0 \text{ より、} C = 0$$

$$\frac{dz}{dt} = v_z(t) = -gt, \quad z(t) = -\frac{1}{2}gt^2 + D$$

$$z(0) = h \text{ より、} D = h \text{ 従って、}$$

$$z(t) = -\frac{1}{2}gt^2 + h$$

力が座標に依存しない場合は単に2回積分すればよい
2階の微分方程式を解くには2つの初期条件が必要⁴

運動方程式の解(2)

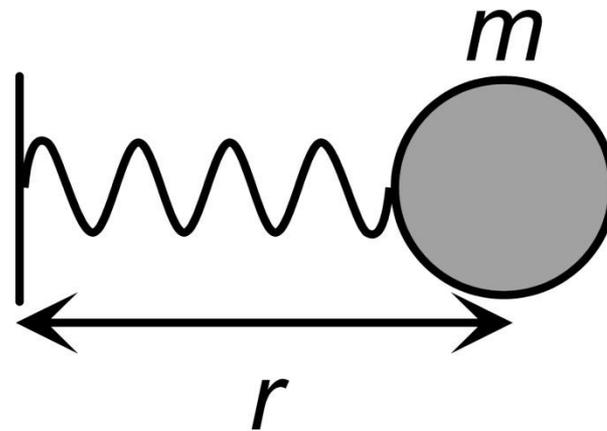
- ばねにつながった玉の運動

質量: m

ばねの長さ: r

ばねの平衡長: r_0

ばね係数: k



$$F = -k(r - r_0)$$

$$m \frac{d^2 r}{dt^2} = -k(r - r_0)$$

変数変換

$$\xrightarrow{q = r - r_0}$$

$$F = -kq$$

$$m \frac{d^2 q}{dt^2} = -kq$$

運動方程式の解(3)

$$\frac{d^2 q}{dt^2} = -\frac{k}{m} q = -\omega^2 q, \quad \omega = \sqrt{\frac{k}{m}} \text{ の一般解は、}$$

$$q(t) = A \cos \omega t + B \sin \omega t$$

$$q(0) = q_0, \quad dq/dt|_{t=0} = v(0) = 0 \text{ とすると、}$$

$$A = q_0, \quad B = 0 \text{ より、}$$

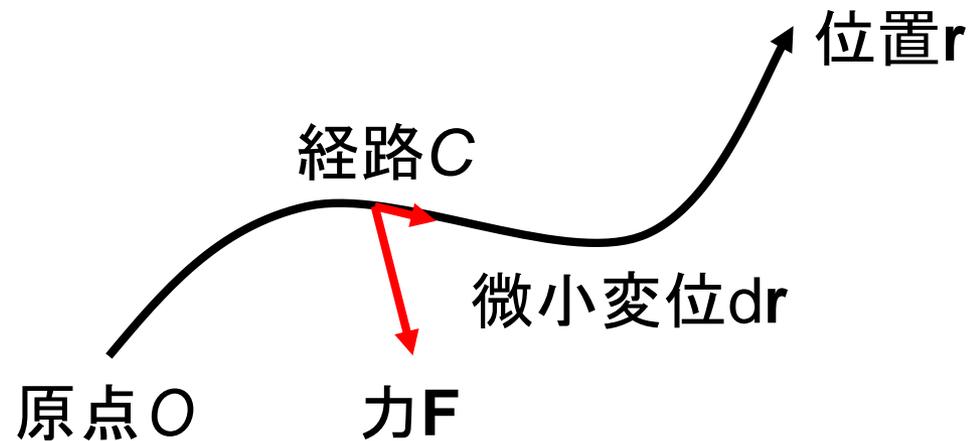
$$q(t) = q_0 \cos \omega t, \quad r = q_0 \cos \omega t + r_0$$

ポテンシャルエネルギーと力(1)

位置 \mathbf{r} におけるポテンシャルエネルギー $E(\mathbf{r})$ の定義

$$E(\mathbf{r}) = -\int_C \mathbf{F} \cdot d\mathbf{r}$$

原点と位置 \mathbf{r} を
結ぶ経路 C に
沿って、力 \mathbf{F} と
微小変位 $d\mathbf{r}$ の
内積を積分



ポテンシャルエネルギーと力(2)

位置 \mathbf{r} から微小変位 $\Delta\mathbf{r}$ させた時のポテンシャルエネルギー変化

$$\begin{aligned} E(\mathbf{r} + \Delta\mathbf{r}) - E(\mathbf{r}) &= E(x + \Delta x, y + \Delta y, z + \Delta z) - E(x, y, z) \\ &= E(x + \Delta x, y + \Delta y, z + \Delta z) - E(x, y + \Delta y, z + \Delta z) \\ &\quad + E(x, y + \Delta y, z + \Delta z) - E(x, y, z + \Delta z) \\ &\quad + E(x, y, z + \Delta z) - E(x, y, z) \\ &= \frac{\partial E}{\partial x} \Delta x + \frac{\partial E}{\partial y} \Delta y + \frac{\partial E}{\partial z} \Delta z = \nabla E \cdot \Delta\mathbf{r} \end{aligned}$$

一方定義より、

$$E(\mathbf{r} + \Delta\mathbf{r}) - E(\mathbf{r}) = -\mathbf{F} \cdot \Delta\mathbf{r}$$

であるから、両者を比較して以下を得る

$$\mathbf{F} = -\nabla E$$

ポテンシャルエネルギー関数が与えられれば力を計算できる

エネルギー保存則(1)

- 孤立系(外界とエネルギーのやりとりがない)ではポテンシャルエネルギー E と運動エネルギー K の和 H が保存する(値が変化しない)

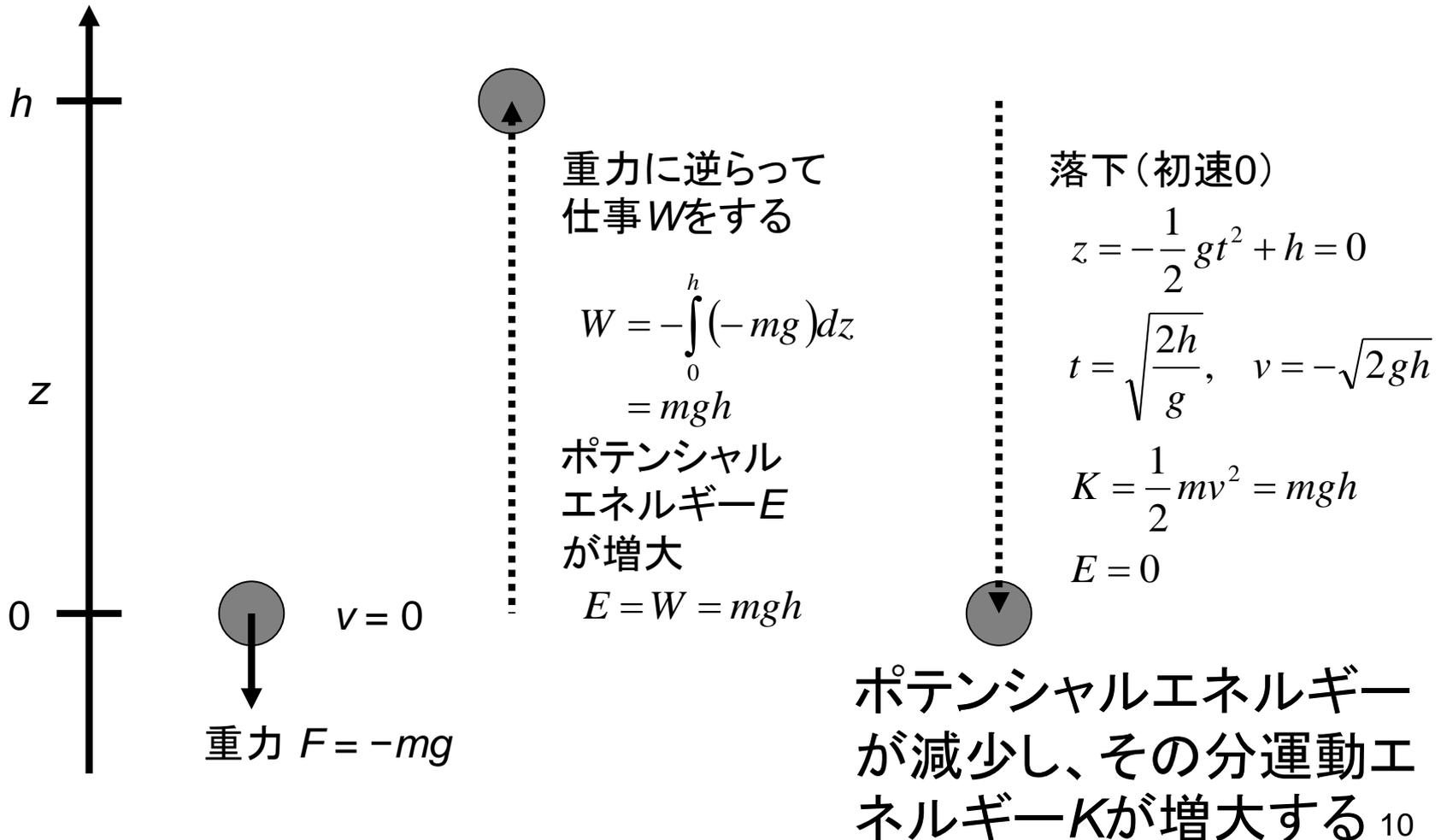
- 運動エネルギー

$$K = \frac{1}{2} m |\mathbf{v}|^2$$

- 保存則の証明

$$\begin{aligned} \frac{dH}{dt} &= \frac{dK}{dt} + \frac{dE}{dt} = m\mathbf{v} \cdot \frac{d\mathbf{v}}{dt} + \frac{d\mathbf{x}}{dt} \cdot \frac{\partial E}{\partial \mathbf{x}} \\ &= m\mathbf{v} \cdot \mathbf{a} - \mathbf{v} \cdot \mathbf{F} = \mathbf{v} \cdot (m\mathbf{a} - \mathbf{F}) = 0 \end{aligned}$$

エネルギー保存則(2)



エネルギー保存則(3)

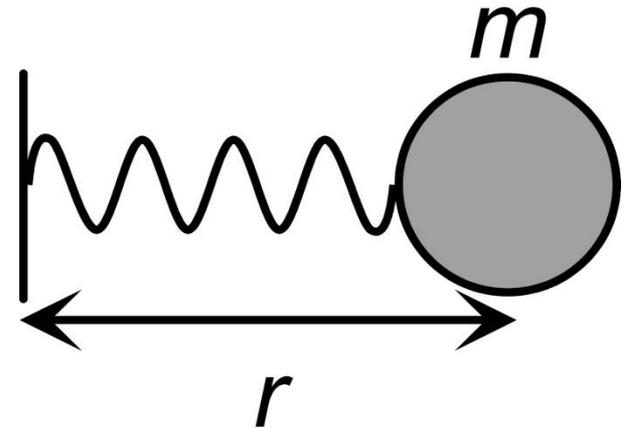
ばねの運動方程式の解

$$r(t) = q_0 \cos \omega t + r_0$$

$$v(t) = dr(t)/dt = -q_0 \omega \sin \omega t$$

$$E(r) = -\int_{r_0}^r -k(r' - r_0) dr' = \frac{k}{2} (r - r_0)^2$$

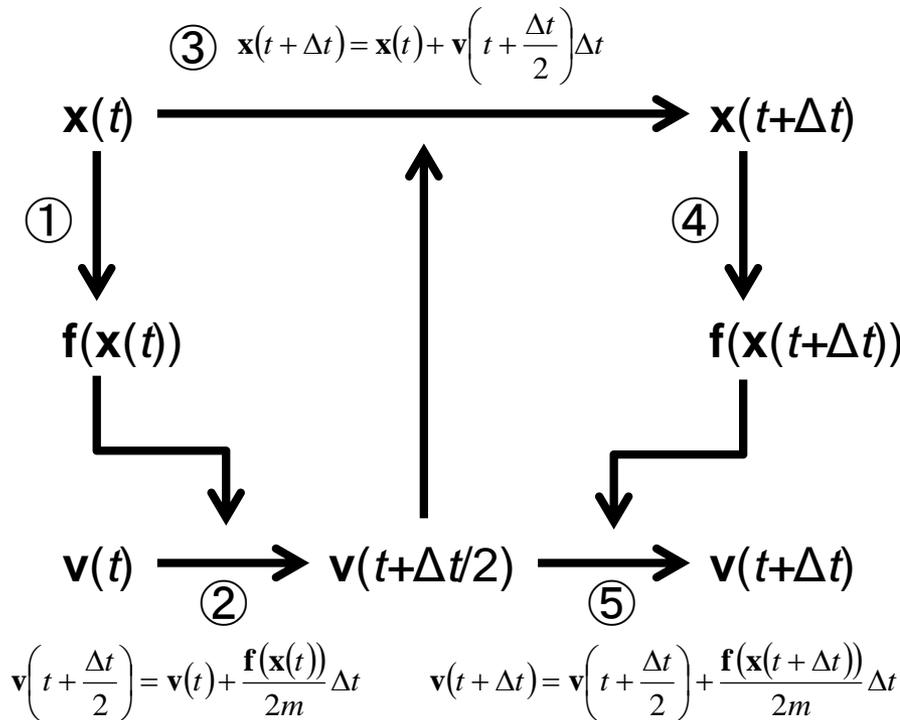
$$\begin{aligned} H &= \frac{m}{2} [v(t)]^2 + \frac{k}{2} [r(t) - r_0]^2 \\ &= \frac{mq_0^2 \omega^2}{2} \sin^2 \omega t + \frac{kq_0^2}{2} \cos^2 \omega t = \frac{1}{2} kq_0^2 \end{aligned}$$



分子動力学法

- 分子動力学 (molecular dynamics, MD) 法では、ポテンシャルエネルギー関数から各原子に働く力を求め、ニュートンの運動方程式に従って、各原子の位置や速度の時間変化を計算する
- 運動方程式は3原子以上の系では解析的に解くことができない→数値解法の利用
- 数値解法では短い時間刻み Δt で逐次的に各原子の位置や速度を求める
- エネルギーがどの程度保存するかを基準に、数値解法の精度を評価する

数値解法 (velocity Verlet法)



- ① 現在の座標から力を計算する
- ② この力を用いて $\Delta t/2$ 後の速度を計算する
- ③ この速度を用いて Δt 後の座標を計算する
- ④ この座標から力を計算する
- ⑤ この力を用いて Δt 後の速度を計算する
- ⑥ ②に戻って目的の時間に達するまで繰り返す

調和振動子のシミュレーション(1)

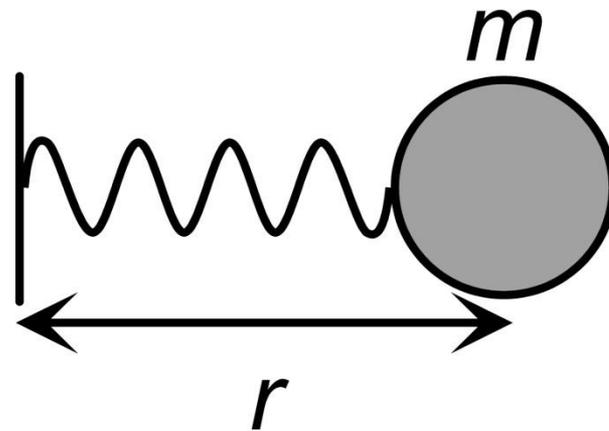
- ばねにつながった玉の運動

質量: m

ばねの長さ: r

ばねの平衡長: r_0

ばね係数: k



$$F = -k(r - r_0)$$

$$m \frac{d^2 r}{dt^2} = -k(r - r_0)$$

変数変換

$$\xrightarrow{q = r - r_0}$$

$$F = -kq$$

$$m \frac{d^2 q}{dt^2} = -kq$$

調和振動子のシミュレーション(2)

- 運動方程式の解析解
初期位置を q_0 、初期速度を0とすると、
$$q(t) = q_0 \cos \omega t, \quad v(t) = -q_0 \omega \sin \omega t$$
- Velocity Verletのperlプログラム (osc.pl)

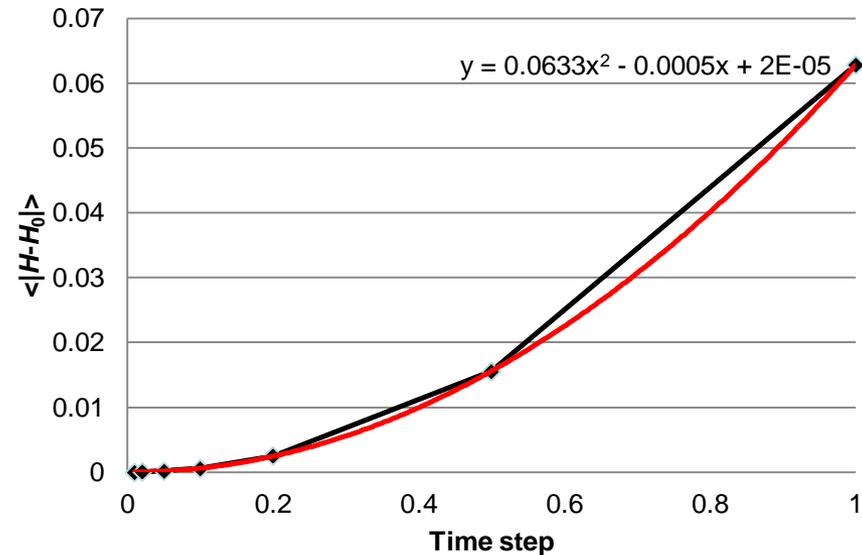
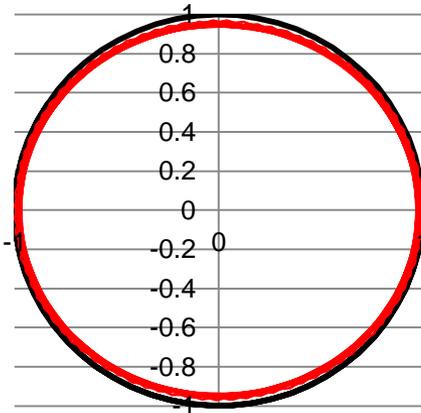
```
$q=1.0;$v=0.0;
$m=1.0;$k=1.0;
$dt=0.01;$nstep=100;
sub calc_force {
  my $q=$_[0];
  my $f=-$k*$q;
  my $e=0.5*$k*$q**2;
  return ($e,$f);
}
open(OUT,">osc$dt.csv");
```

```
($e,$f)=calc_force($q);
for($i=1;$i<=$nstep;$i++) {
  $v+=0.5*$f/$m*$dt;
  $q+=$v*$dt;
  ($e,$f)=calc_force($q);
  $v+=0.5*$f/$m*$dt;
  $H=0.5*$m*$v**2+$e;
  print OUT $i*$dt,",",",$q,
    ",",",$v,",",",$H,"¥n";
}
```

課題1

- 調和振動子のシミュレーションプログラム
osc.plを用いて、 Δt を0.01、0.02、0.05、
0.1、0.2、0.5、1と変化させたときの全エネルギー(H)の初期値($H_0 = 0.5$)との差の絶対値($\langle |H - H_0| \rangle$)の平均を求めよ
- 時間刻みを横軸、 $\langle |H - H_0| \rangle$ を縦軸としてプロットせよ(osc.xlsxを使用しても良い)
- 時間刻みが大きくなると、 $\langle |H - H_0| \rangle$ はどのように変化するか考察せよ

時間刻みと誤差の関係



$\$dt$ (時間刻み) を 0.1 (黒)、0.5 (赤) としたときの、各時刻における ($\$q$, $\$v$) の組をプロット

$\$dt$ を 0.01、0.02、0.05、0.1、0.2、0.5、1 と変化させたときの全エネルギー ($\$H$) の初期値 ($H_0 = 0.5$) との差の絶対値 ($\langle |H - H_0| \rangle$) を、時間刻みを横軸、 $\langle |H - H_0| \rangle$ を縦軸としてプロット

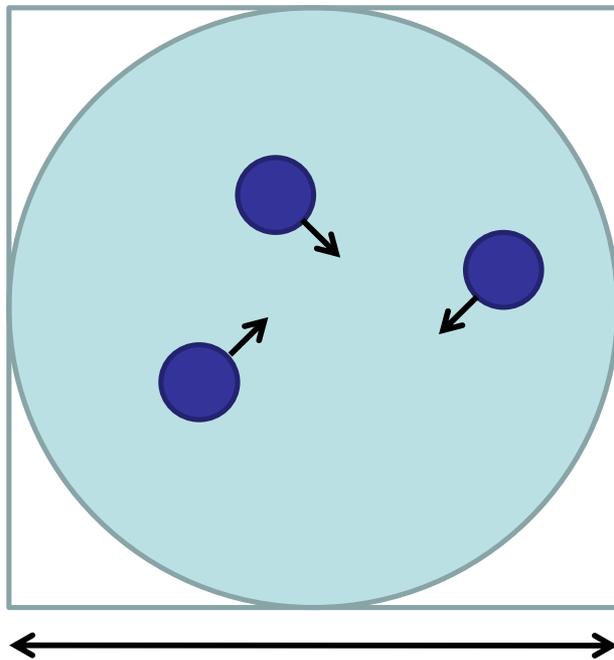
$$H = \frac{mv^2}{2} + \frac{kq^2}{2} = 0.5 \quad \blacktriangleright \quad v^2 + q^2 = 1 \quad |H - H_0| \text{ は円からのずれに対応} \quad 17$$

適切な時間刻みの選択

- 時間刻みを小さくすると、全エネルギーのずれは小さくなる
- 一般的に、最も速い運動の周期の10分の1から20分の1程度の時間刻みがいられる
- タンパク質の場合、最も速い運動は、X-H伸縮運動 (X=C、N、O、S) で、 3000 cm^{-1} 程度 (周期は約10 fs)
- 従って、時間刻み Δt には0.5~1.0 fsを用いる

多粒子系のシミュレーション(1)

- Van der Waals相互作用する多粒子系のシミュレーション(vdw.pl)



```
$natom=3;           # Number of particles
$width=10.0;        # Width of initial
                    # particle distribution
$scale=1.0;         # Scaling factor for
                    # initial velocity
$fcap=1.0;          # Force constant for
                    # spherical boundary
$sigma=1.0;         # Atom radius
$epsilon=1.0;       # Well depth
$mass=1.0;          # Atomic mass
$nstep=100000;     # Number of MD steps
$nsave=100;         # Frequency of saving
                    # trajectory
$dt=0.001;          # Time step
$seed=190523;       # Random seed
```

\$width

多粒子系のシミュレーション(2)

- 初期配置
 - 原点を中心とした、一辺が\$widthの立方体の中にランダムに配置
- 初期速度
 - ランダム(大きさは\$scaleで変更可)
- ポテンシャルエネルギー関数

$$E = \sum_{i=1}^N \sum_{j=i+1}^N 4\epsilon \left(\frac{\sigma^{12}}{r_{ij}^{12}} - \frac{\sigma^6}{r_{ij}^6} \right) + \sum_{i=1}^N E_{\text{cap}}(r_i)$$

$$E_{\text{cap}}(r_i) = \begin{cases} 0 & r_i < r_{\text{cut}} \\ f_{\text{cap}}(r_i - r_{\text{cut}})^2 & r_i \geq r_{\text{cut}} \end{cases}$$

r_i は原点からの距離
 r_{cut} は\$widthの1/2

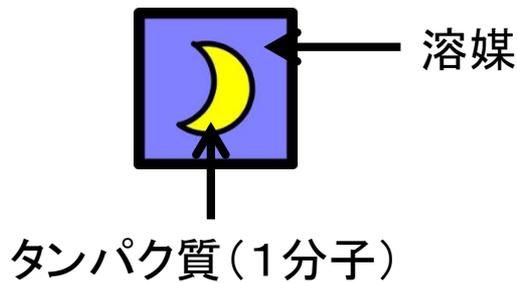
多粒子系のシミュレーション(3)

- 計算結果はUCSF Chimeraで可視化できる
- Chimeraのアイコンをダブルクリック
- メニューの「Tools」→「MD/Ensemble Analysis」→「MD Movie」を選択し、Trajectory formatを「PDB」、PDB frames contained inを「Single file」とし、ファイルに生成された「vdw.pdb」を指定して「OK」
- 再生ボタンをクリックしてアニメーションを開始
- パラメータを変えて運動の変化を観察せよ

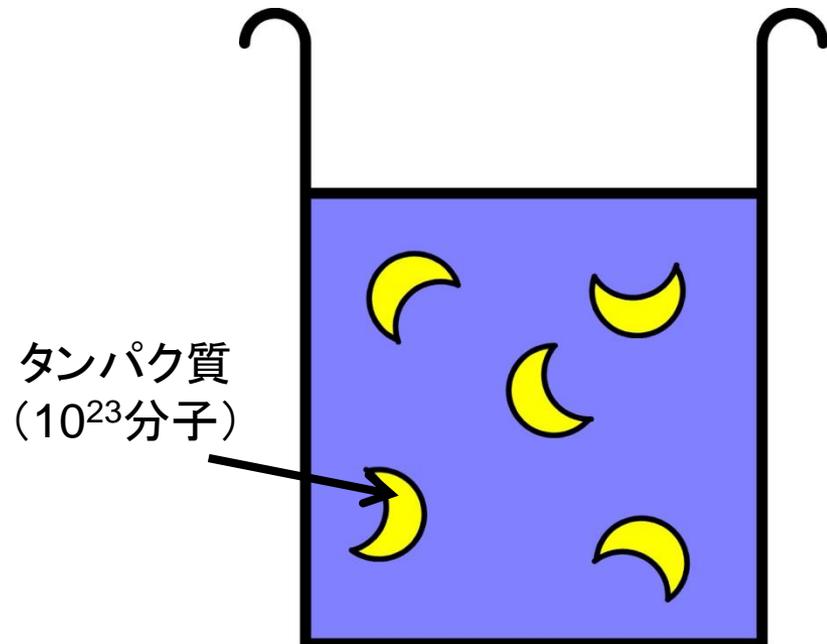


実験データとの比較(1)

シミュレーション
のシステム

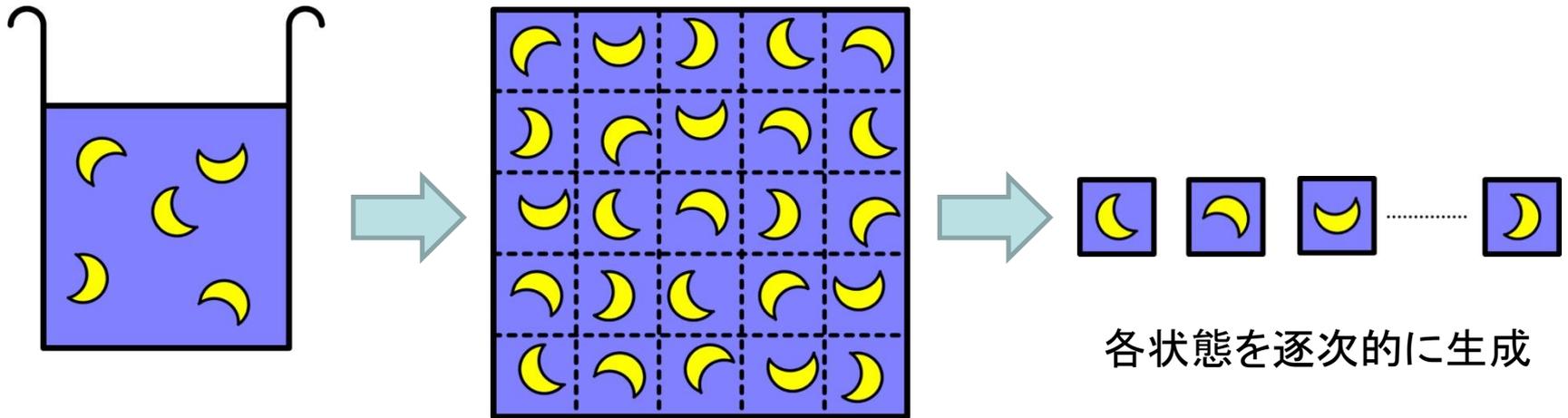


現実のシステム



- 分子シミュレーションで得られる結果は、実験データと比較できるのだろうか？

実験データとの比較(2)



熱だけを交換できる仕切りで分割

各状態を逐次的に生成

- 現実のシステムに存在する状態を、分子シミュレーションによって逐次的に生成することで、現実のシステムを再現する

実験データとの比較(3)

- 実験データは、全系を構成する各系(タンパク質とその周辺環境)の状態によって決まる値の平均値(状態の出現確率での重みつき平均)

$$\langle A \rangle = A_i \rho_i, \quad \sum_i \rho_i = 1 \quad \rho_i: \text{状態 } i \text{ の出現確率}$$

- 各状態の出現確率は**カノニカル分布**に従う

$$\rho_i = Z^{-1} \exp(-e_i/k_B T) \quad e_i: \text{状態 } i \text{ のエネルギー}$$

$$Z = \sum_i \exp(-e_i/k_B T) \quad Z: \text{分配関数}$$

→**カノニカル分布**に従うように状態を生成すれば良い

調和振動子の場合(1)

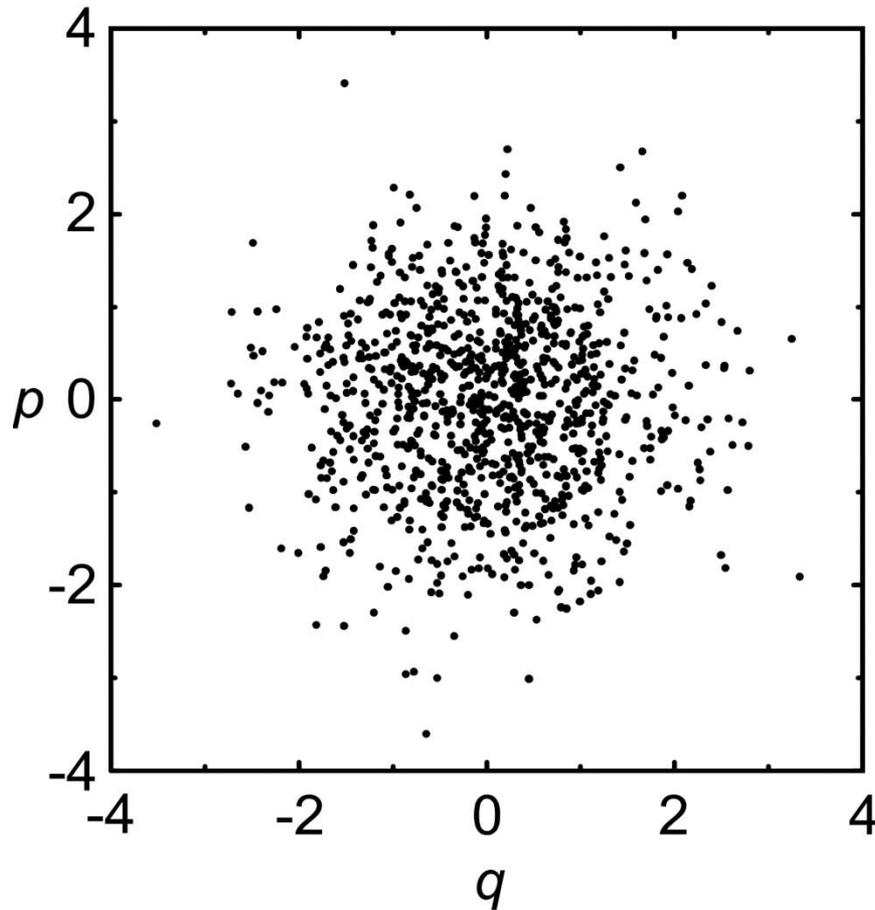
- 調和振動子の状態(q, p)のカノニカル分布

$$H(q, p) = \frac{p^2}{2m} + \frac{k}{2}q^2, \quad \rho(q, p) = \frac{\exp\left[-\frac{H(q, p)}{k_B T}\right]}{\int \exp\left[-\frac{H(q, p)}{k_B T}\right] dq dp}$$

- この分布に従うように状態(q, p)を生成する
 - 全エネルギーの平均値を用いて精度を評価する

$$\langle H \rangle = \int H(q, p) \rho(q, p) dq dp = k_B T$$

調和振動子の場合(2)



- $m = k = k_B T = 1$ の時、この分布は分散1の正規分布と一致

$$\rho(q, p) = \frac{1}{2\pi} \exp\left(-\frac{q^2 + p^2}{2}\right)$$

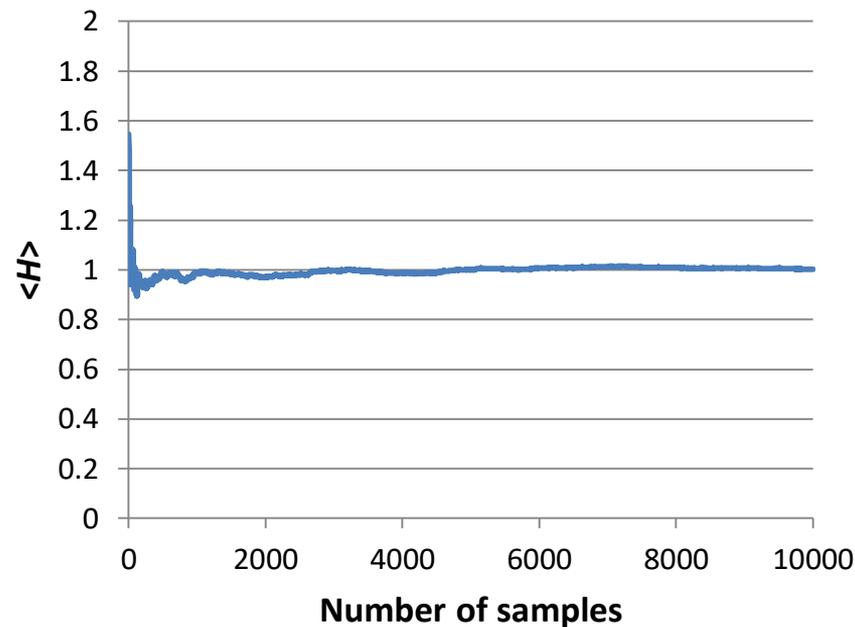
- 発生させる状態の数を増やしていき、 $\langle H \rangle$ の収束を見る

プログラム例

```
$kT=1.0;
$pi=atan2(1.0,1.0)*4.0;
$npt=10000;

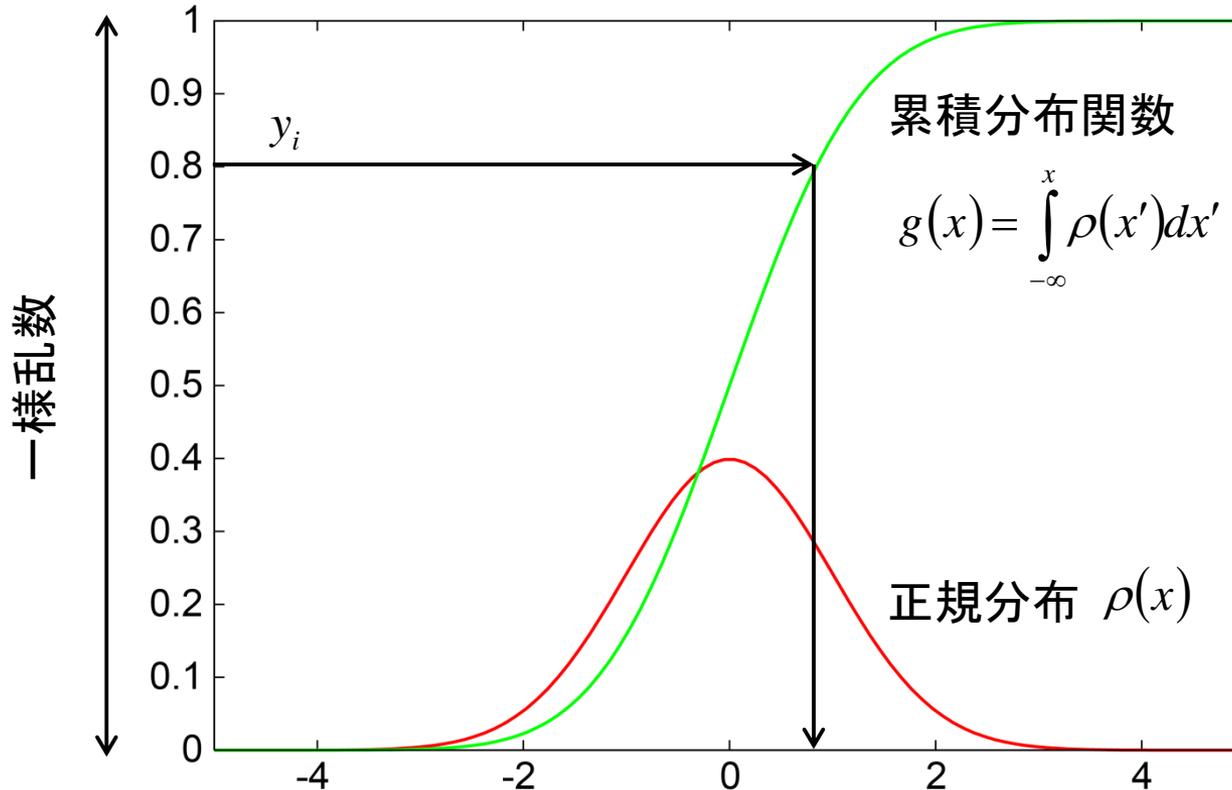
open(OUT,">normal.csv");
$val1=0.0;
for($i=1;$i<=$npt;$i++) {
    $x1=rand;
    $x2=rand;
    #一様乱数を正規分布に変換
    $q=sqrt(-2.0*$kT*log($x1))*cos(2.0*$pi*$x2);
    $p=sqrt(-2.0*$kT*log($x1))*sin(2.0*$pi*$x2);
    $H=0.5*$q**2+0.5*$p**2;
    $val1+=$H;           #標本点における全エネルギーを平均
    if($i % 100 == 0) {
        printf(OUT "%d,%f¥n",$i,$val1/$i);
    }
}
```

調和振動子の場合(3)



- 理論値 $\langle H \rangle = 1$ に素早く収束することがわかる

標本点の発生法(1)



正規分布に従う乱数に変換

$$g(x_i) = y_i, \quad x_i = g^{-1}(y_i)$$

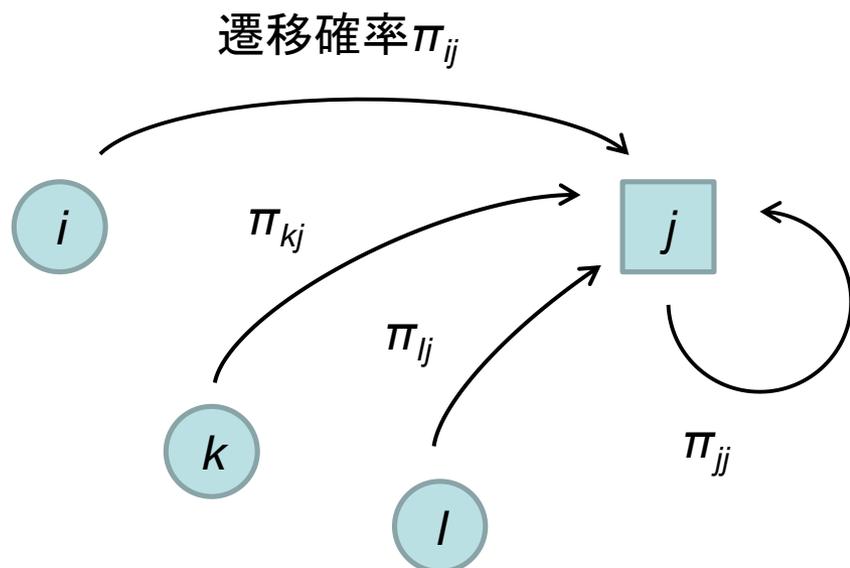
標本点の発生法(2)

- 前出の方法は累積分布関数が計算可能な場合のみ適用可能
- 生体高分子を含む系の場合、共有結合相互作用と非共有結合相互作用の関係が複雑なので、累積分布関数を解析的に計算できない
- また、内部自由度が大きいため数値的にも計算できない



Markov連鎖を利用する

Markov連鎖



- 状態 i から状態 j に遷移する確率を π_{ij} とする
- 遷移前の状態 i の存在確率を ρ_i^0 とすると、遷移後の状態 j の存在確率は、以下で表される

$$\rho_j^1 = \sum_i \rho_i^0 \pi_{ij}$$

- また、遷移確率は以下を満たす

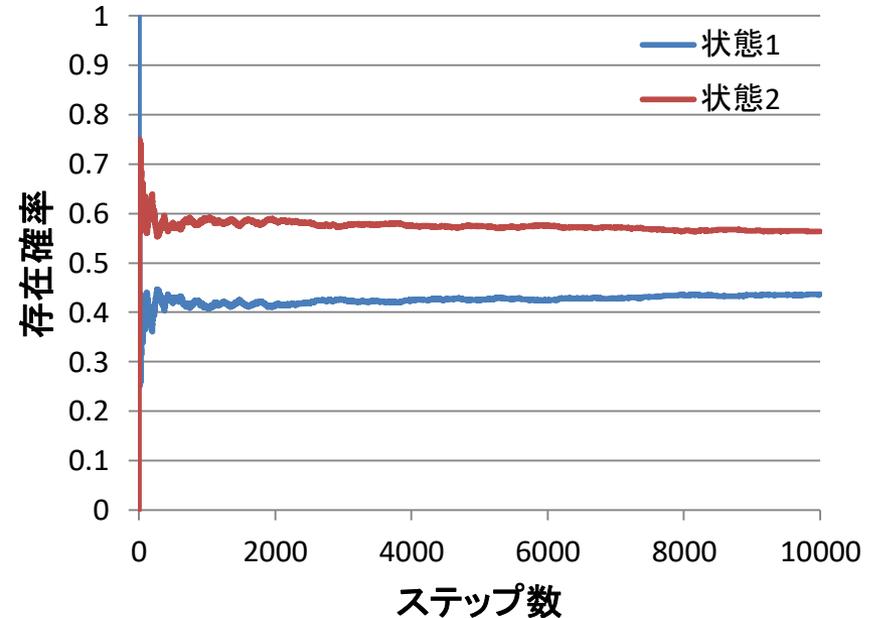
$$\sum_j \pi_{ij} = 1$$

Markov連鎖の例(1)

- 状態が2つと仮定
- 遷移確率を以下のように定める

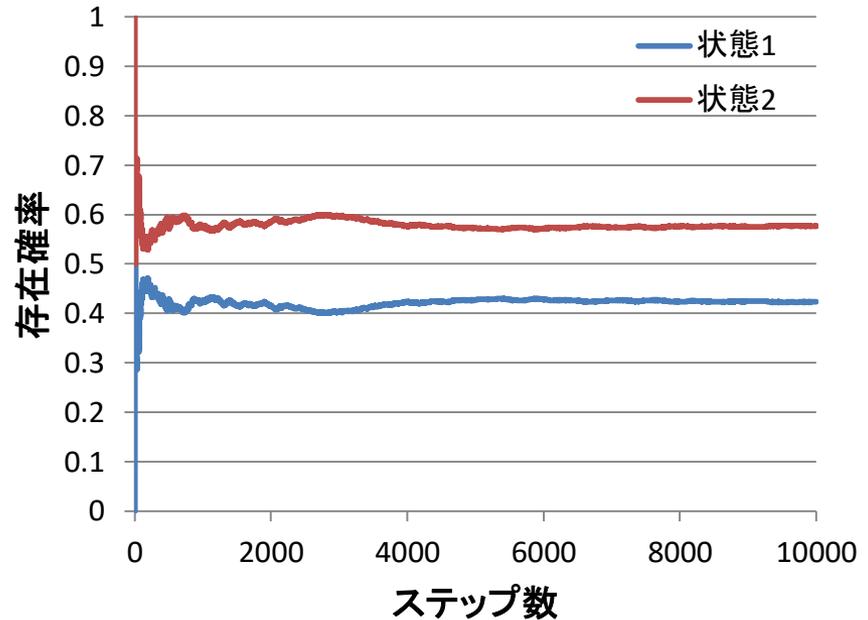
$\pi_{11} = 0.6$	$\pi_{12} = 0.4$
$\pi_{21} = 0.3$	$\pi_{22} = 0.7$

- はじめ状態1にあったとすると、状態1、状態2の存在確率は右図のようになる



Markov連鎖の例(2)

- はじめが状態2
であっても同じ
存在確率に収束



- 収束した段階では方程式 $\rho = \rho\pi$ を満たす
- 存在確率 ρ は、遷移行列 π の固有ベクトル
→ 遷移行列が決まれば一意的に決まる

Metropolisの方法(1)

- 存在確率から逆に遷移行列を求めたい
- 詳細釣り合いが成り立てば $\boldsymbol{\rho} = \boldsymbol{\rho}\boldsymbol{\pi}$ が成り立つ

$$\rho_i \pi_{ij} = \rho_j \pi_{ji} \Rightarrow \sum_i \rho_i \pi_{ij} = \sum_i \rho_j \pi_{ji} = \rho_j \sum_i \pi_{ji} = \rho_j$$

- Metropolisらは詳細釣り合いを満たす遷移行列として以下を提案

$$\begin{cases} \pi_{ij} = \alpha_{ij} & \text{if } \rho_j \geq \rho_i \text{ and } i \neq j \\ \pi_{ij} = \alpha_{ij} (\rho_j / \rho_i) & \text{if } \rho_j < \rho_i \text{ and } i \neq j \end{cases}$$

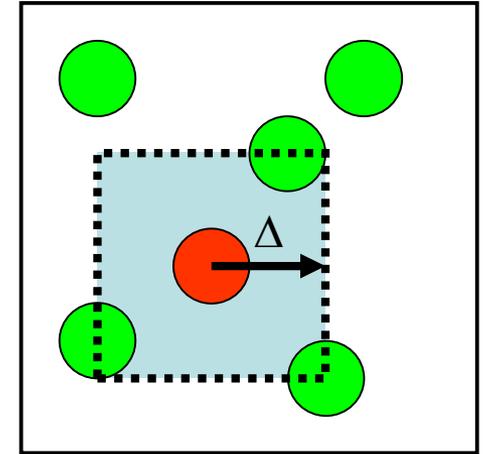
$$\alpha_{ij} = \alpha_{ji}, \quad \sum_{j \neq i} \alpha_{ij} = 1, \quad \pi_{ii} = 1 - \sum_{j \neq i} \pi_{ij}$$

Metropolisの方法(2)

1. 注目している原子を、その原子を中心とした、1辺 2Δ の立方体の範囲内でランダムに遷移

$$\alpha_{ij} = \frac{1}{N_{\Delta}} \quad \text{遷移先が立方体の中}$$

$$\alpha_{ij} = 0 \quad \text{遷移先が立方体の外}$$



2. 遷移先のエネルギー e_j が、もとの状態のエネルギー e_i よりも小さければ、その遷移を採択し、大きければ、以下の確率で採択

$$\rho_j / \rho_i = \exp[-(e_j - e_i) / k_B T] = \exp(-\Delta e_{ji} / k_B T)$$

3. 採択されなかった場合は、もとの状態を新しい状態とする

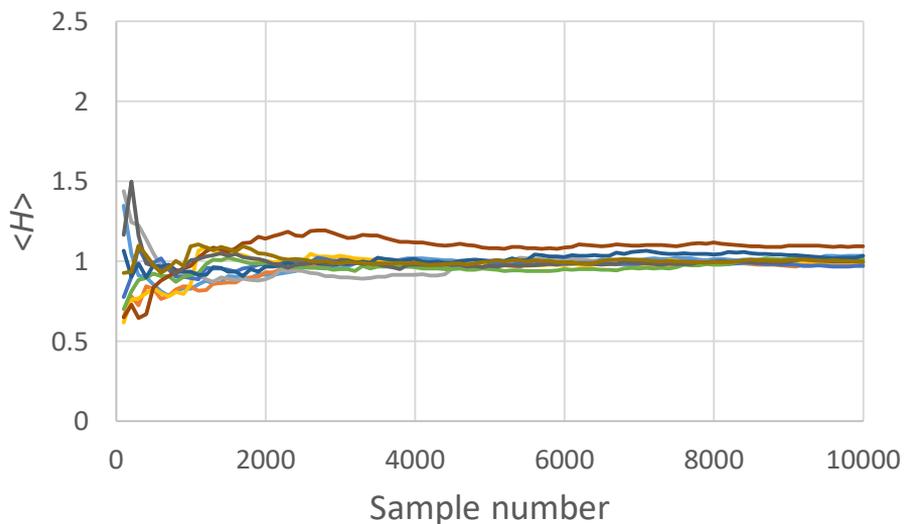
$$\pi_{ii} = 1 - \sum_{j \neq i} \pi_{ij} = \sum_{j \neq i} (\alpha_{ij} - \pi_{ij}) = \sum_{\substack{j \neq i \text{ and} \\ \rho_j < \rho_i}} \alpha_{ij} (1 - \rho_j / \rho_i)$$

課題2

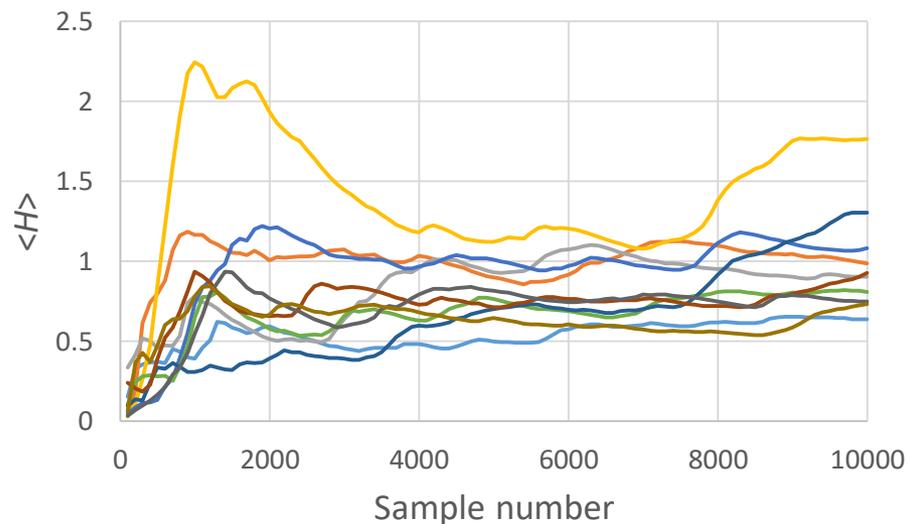
- Metropolisの方法を調和振動子の系に適用する
- 講義のページからmetropolis.plをダウンロードし、アイコンをダブルクリックして実行せよ
 - サンプル数を横軸に、 $\langle H \rangle$ を縦軸にプロットし、 $\langle H \rangle$ が1に収束することを確認せよ
- 収束の仕方が変位の最大値 δ に依存することを確認せよ
 - 特に $\delta=0.1$ としたときにどうなるか確認せよ
- このようになる理由を考察せよ

結果の例

$\$delta = 1.0$



$\$delta = 0.1$



- 出力ファイルには10回の独立なランの結果が含まれる
- $\$delta = 0.1$ についてステップ数($\$nstep$)を大きくしてみよ

生体高分子への適用

- 生体高分子では、「状態」は立体構造に対応
- Metropolisの方法は生体高分子から原子を任意に選び、その位置を動かすことで実現可能
- ただし、原子を動かすと大抵の場合共有結合長を変えることになり、エネルギーが増加する場合が多い
 - 棄却される確率が高い
- このため、共有結合長や共有結合角を固定し、二面角のみを動かすのが一般的だが困難も多い
 - 複数分子の扱いが難しい
 - タンパク質のコアのように密にパックしている部分では、二面角を回転すると原子が衝突する可能性が高い

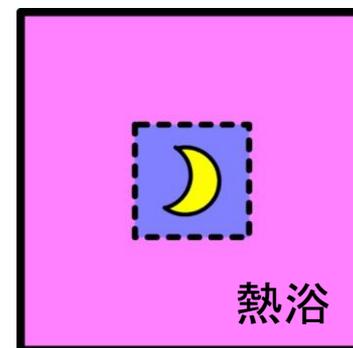
定温分子動力学法(1)

- 分子動力学法において、温度が一定になるように制御することで、canonical分布に従う立体構造分布を生成することができる
 - 温度は速度から計算される $\frac{3}{2}Nk_B T = \sum_{i=1}^N \frac{m_i |\mathbf{v}_i|^2}{2}$
- 直交座標系で扱えるためモンテカルロ法よりも生体高分子に適用しやすい
- 平均値を計算する際には時間平均として計算する

定温分子動力学法(2)

- 能勢の方法
- 能勢-Hoover chain法
 - 能勢の方法の改良版
- 束縛法
 - 座標のみcanonical分布に従う
- Langevin dynamics
 - ランダム力と摩擦により温度を制御
- Berendsenのcoupling法
 - Canonical分布を与えない
 - 簡便で安定なため、よく用いられる

} 系が熱浴と接している
と考え、熱浴に自由度
を与える



Langevin dynamics

- 熱浴の仮想粒子が原子に衝突してエネルギーを交換する
 - 原子が「熱い」時は、熱浴にエネルギーが逃げ、「冷たい」時は、熱浴からエネルギーを得る

- 運動方程式

$$m\mathbf{a} = \mathbf{F}(\mathbf{x}) - \gamma\mathbf{v} + \mathbf{R}(t)$$

摩擦力 仮想粒子との衝突による力

- ランダム力 \mathbf{R} は、以下を満たす

$$\langle \mathbf{R}(t) \rangle = 0, \quad \langle \mathbf{R}(t) \cdot \mathbf{R}(t') \rangle = 6k_B T \gamma \delta(t - t')$$

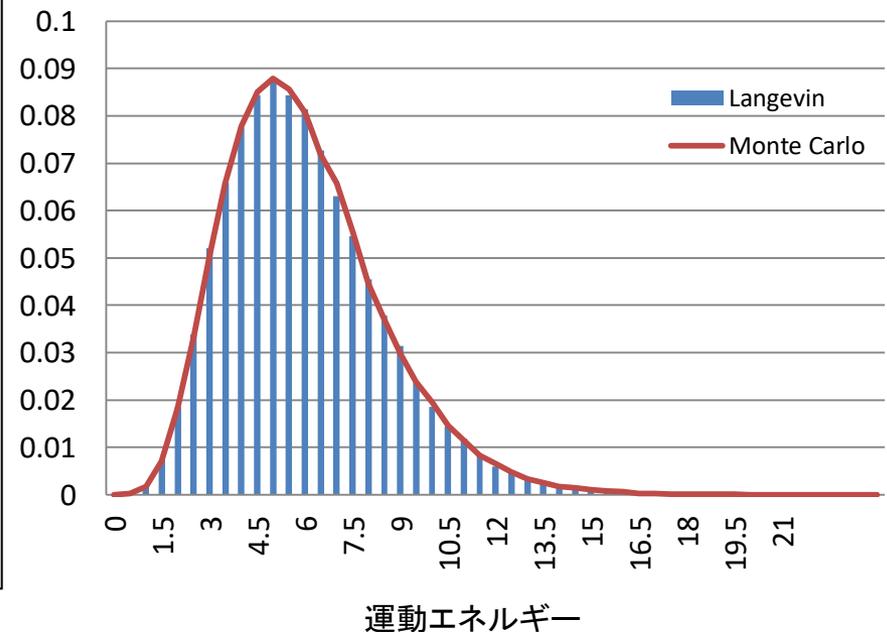
平均 分散・共分散

計算例

- Van der Waals相互作用する多粒子系のシミュレーション (vdw_langevin.pl)

```
$natom=4;           # Number of particles
$width=10.0;        # Width of initial
                    # particle distribution
$fcap=1.0;          # Force constant for
                    # spherical boundary
$sigma=1.0;         # Atom radius
$epsilon=1.0;       # Well depth
$mass=1.0;          # Atomic mass
$nstep=100000;      # Number of MD steps
$nsave=100;         # Frequency of saving
                    # trajectory
$dt=0.001;          # Time step
$seed=190523;       # Random seed
$gamma=10.0;        # friction coefficient
$kT=1.0;
```

運動エネルギーの確率密度分布



参考 : Berendsenのcoupling法

1. Velocity Verlet法の各ステップ終了後に瞬間的な温度 T を計算する

$$\frac{3}{2} Nk_{\text{B}}T' = \sum_{i=1}^N \frac{m_i |\mathbf{v}_i|^2}{2}$$

2. 速度を χ 倍スケールする (定数 τ を大きくすると速度のスケールの程度は小さくなる)

$$\chi = \left[1 + \frac{\Delta t}{\tau} \left(\frac{T}{T'} - 1 \right) \right]^{1/2}$$

課題の提出

- 課題1の結果と考察、課題2の結果と考察を1つのPowerPointファイルにまとめ、寺田宛 tterada@iu.a.u-tokyo.ac.jp に送ること
- その際件名は「分子モデリング課題」とし、本文に氏名と学生証番号、受講生IDを明記すること