10/26/2022

- Quiz 2
- Monte Carlo Integration
- Markov chain Monte Carlo

Monte Carlo

- Today's lecture is intended to help you achieve the following learning objective: Explain the key objectives of molecular simulation

- Monte Carlo Integration
- Markov Chain Monte Carlo

Monte Carlo Integration

If you keep guessing, you may eventually get it

Expectation Values

- Expectation values are $\langle A \rangle = \frac{\sum_j A(x_j)p(x_j)}{\sum_j p(x_j)}$.
  - A(x) is any observable. For molecular systems, useful observables include,
    - distance between two atoms (e.g. hydrogen bond)
    - collective variable, e.g. angle between three domains
    - probability that a collective variable is within a bin
    - energies or moments of energy, yielding free energy differences
  - p(x) is a normalized probability density, e.g. uniform
  - x_j is a vector, indexed by j
  - the sum is over all possible states
- For a continuous state space, e.g. molecular mechanics, $\langle A \rangle = \int A(x)p(x)dx$.

Calculating Expectation Values

- Two general approaches to calculating $\langle A \rangle = \dfrac{\sum_j A(x_j)p(x_j)}{\sum_j p(x_j)}$
  - Enumeration
    - compute $A(x_j)$ and $p(x_j)$ for all possible x_j; weighted average.
    - for continuous spaces, integrals may be approximated by quadrature, e.g. Simpson's rule
  - Monte Carlo Integration (named after famous gambling city)
    - draw (pseudo-)random samples x_j with probability $p(x_j)$; estimate with simple average
    - estimate converges to true value with many samples

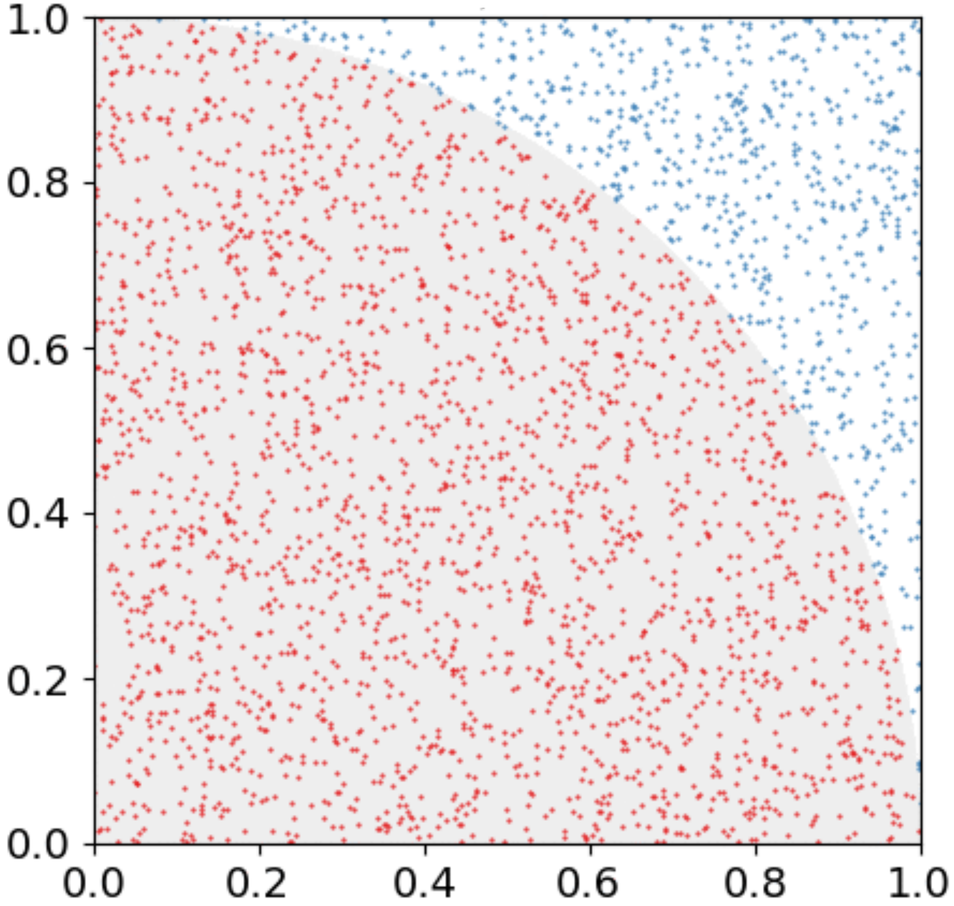
Example: Dice

- How do you compute the expectation value of the sum of two six-sided dice?
  - (2 + 3*2 + 4*3 + 5*4 + 6*5 + 7*6 + 8*5 + 9*4 + 10*3 + 11*2 + 12)/36 = 7
  - Roll the dice many times and calculate the average

	1	2	3	4	5	6
1	2	3	4	5	6	7
2	3	4	5	6	7	8
3	4	5	6	7	8	9
4	5	6	7	8	9	10
5	6	7	8	9	10	11
6	7	8	9	10	11	12

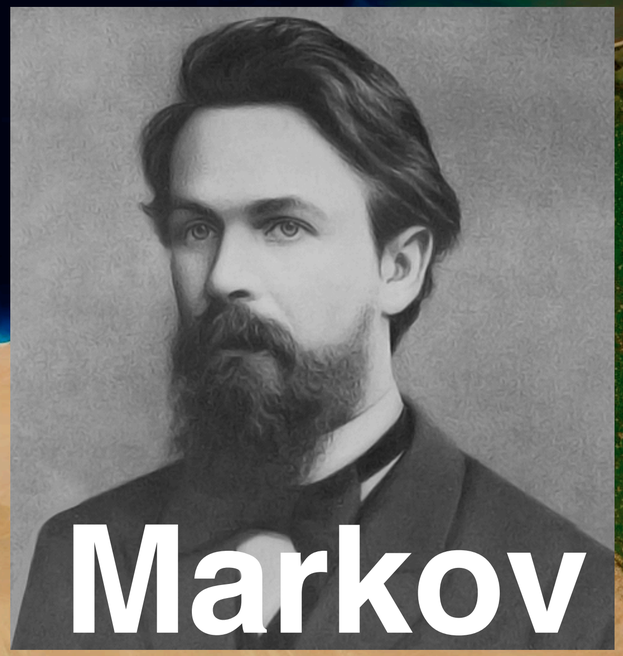
Example: Calculating π

- Sample points from a two-dimensional uniform distribution between 0 and 1
- Estimate $\langle H(d(x) - 1) \rangle$
  - $d(z)$ is the distance from the origin
  - $H(z)$ = 1 if z < 0
  - $H(z)$ = 0 otherwise
  - the fraction of points inside square that are inside the circle; thus, $A = \langle H(d(x) - 1) \rangle \, r^2$
- Since $A = \frac{1}{4}\pi r^2$ for a quarter circle, $\pi = 4 \langle H(d(x) - 1) \rangle$



Limitations

- Both examples were for
  - a simple distribution (uniform)
  - low dimensions
- Drawing independent random samples is difficult for complex high-dimensional probability distributions, such as biological macromolecules



Markov Chain Monte Carlo

A smarter way to sample the Nile

Sampling the Nile

- Let's say that you want to calculate the average depth of the Nile. If your boat is on the Nile river, which is the smarter way to sample other places in the Nile?
    A. Throw darts at a map of Egypt. Check whether the darts landed in the Nile. If so, move there and measure the depth.
    B. Randomly move within the river. Measure at regular time intervals.

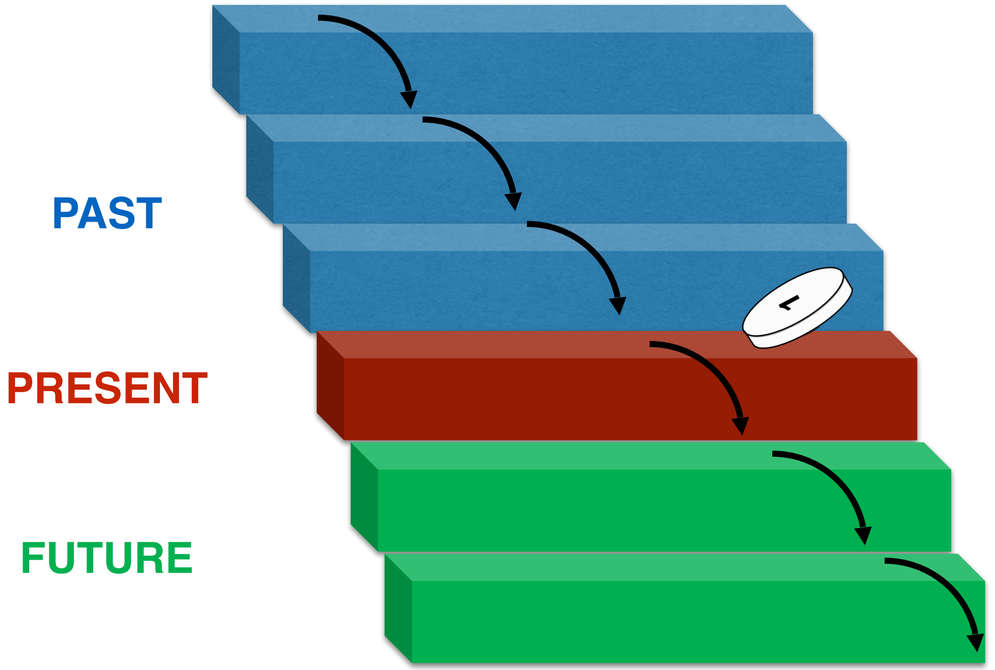
- In the former, samples would be <u>independent and identically distributed</u> (i.i.d.)
- In the latter, samples would be from a <u>stochastic process,</u> a sequence of indexed random variables

$\boldsymbol{X}$ = stochastic process $\quad \boldsymbol{X} = \{X_0, X_1, X_2 \ldots\}, \;\; X_i \in \chi,$

χ = state space

Markov Chains

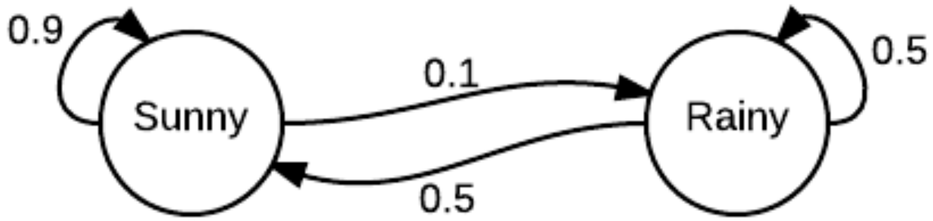
- Special class of stochastic process
- Markovian = memoryless
  - future depends only on the present, not the past
  - independent and identically distributed (i.i.d.) is a special case
- But wait… can you use samples from a Markov chain to estimate expectation values?
  - Yes, if it satisfies certain conditions which I will explain



$$P\left(X_{i+1} \middle| X_i = x_i, X_{i-1} = x_{i-1} \ldots, X_0 = x_0\right) = P\left(X_{i+1} \middle| X_i = x_i\right)$$

Representing Transition Probabilities

- Graphs
  - nodes are states
  - edges are transition probabilities



- Matrices
  - with $Pr(j \mid i) = P_{i,j}$.
  - Transition matrices satisfy

$$\sum_{j=1}^{\alpha} P_{i,j} = 1$$

transition matrix

$$Q = \begin{pmatrix} 0.9 & 0.1 \\ 0.5 & 0.5 \end{pmatrix}$$

probability vector evolution

$$\boldsymbol{\pi}_{k+1} = \boldsymbol{\pi}_k Q$$

If it is sunny on Monday, what is the probability that it is sunny on Tuesday?

Stationary Probability

transition matrix

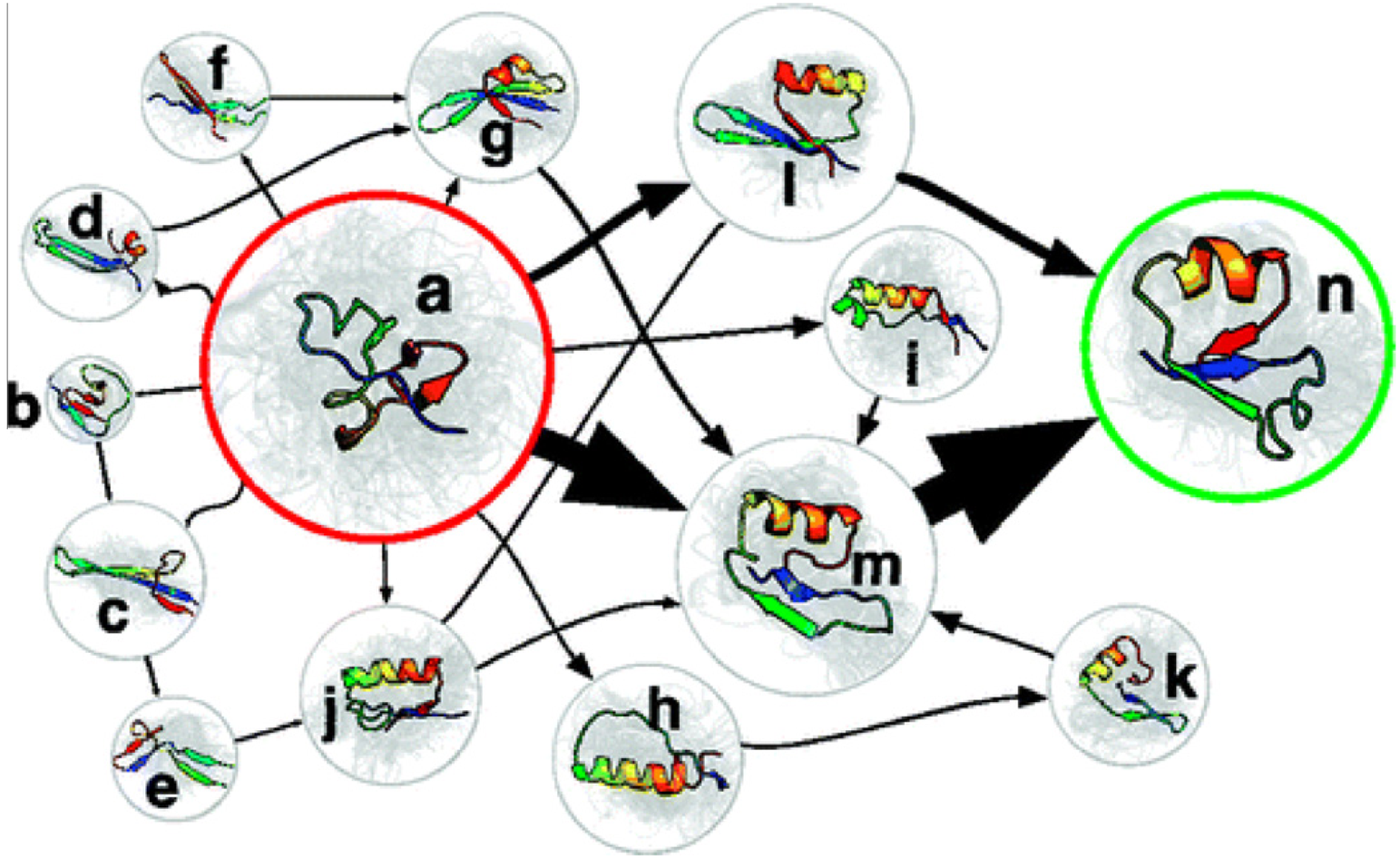
$$Q = \begin{pmatrix} 0.9 & 0.1 \\ 0.5 & 0.5 \end{pmatrix}$$

probability vector evolution

$$\boldsymbol{\pi}_{k+1} = \boldsymbol{\pi}_k Q$$

- If the probability of Monday being sunny is 1/2, what is the probability of it being sunny on Tuesday?
  - 1/2*0.9 + 1/2*0.5 = 0.7
- If the probability of Monday being sunny is 5/6, what is the probability of it being sunny on Tuesday?
  - 5/6*0.9 + 1/6*0.5 = (4.5 + 0.5)/6 = 5/6
- This is known as the stationary probability, which does not change under application of the transition matrix

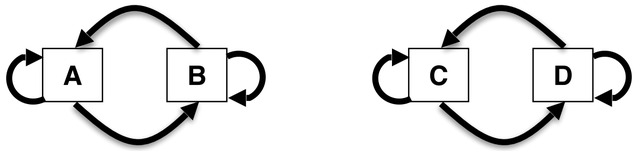
Markov State Model



. V.A. Voelz et al., J. Am. Chem. Soc. 132 (5) (2010) 1526–1528.

Expectations from Stochastic Processes

- Samples from a Markov chain can be used to estimate expectation values if
  - it is irreducible / ergodic - simulations starting at one state will visit all other states
  - it is aperiodic
  - the stationary distribution is the target distribution $\pi(\cdot)$ (e.g. Boltzmann)
- Distribution of $\{X_n\}$ converges to $\pi(\cdot)$; expectations based on average
- For more details, see Minh, Minh, and Nguyen (2012)



Can you use a Markov chain with this kernel to estimate expectations?

Biased Estimation

- Estimates are biased unless
  - initial value from target distribution
  - the chain is regenerative (returns to initial state) and expectation based on data between regeneration times. See Minh, Minh, and Nguyen (2012) for more details.
- Bias diminishes with sample size

References

- Grossfield A, Zuckerman DM. Quantifying uncertainty and sampling quality in biomolecular simulations. Annu Rep Comput Chem. 2009 Jan 1;5:23-48. doi: 10.1016/S1574-1400(09)00502-7. PMID: 20454547; PMCID: PMC2865156.
- Kalos, Malvin H.; Whitlock, Paula A. (2008). Monte Carlo Methods.
- Levin DA, Peres Y and Wilmer EL, “Markov Chains and Mixing Times”, American Mathematical Soc., Oct 31, 2017, ISBN: 1470429624, 9781470429621
- Minh, D. L.; Minh, D. D. L.; Nguyen, A. L. Regenerative Markov Chain Monte Carlo for Any Distribution. Communications in Statistics - Simulation and Computation 2012, 41 (9), 1745–1760. https://doi.org/10.1080/03610918.2011.615433.