REFERENCES

1. R. Chen and J. Liu. Sequential monte carlo methods for dynamic systems. *Journal of American Statistical Association*, 93:1032–1043, 1998.

2. M. O. Dayhoff, R. M. Schwartz, and B. C. Orcutt. *Atlas of Protein Sequence and Structure*. National Biomedical Research Fundation, Washington, D.C, 1978.

3. M.O. Dayhoff, R.M Schwartz, and B.C Orcutt. *Atlas of Protein Sequence and Structure, vol.5, suppl. 3.*, chapter A model of evolutionary change in proteins, pages pp. 345–352. National Biomedical Research Foundation, Washington, D.C, 1978.

4. A. Doucet, N. De Freitas, Gordon N., and A. Smith. *Sequential Monte Carlo Methods in Practice*. Springer Verlag, 2001.

5. David J. Earl and Michael W. Deem. Parallel tempering: Theory, applications, and new perspectives. *Phys. Chem. Chem. Phys.*, 7:3910–3916, 2005.

6. J. Felsenstein. Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.*, 17:368–376, 1981.

7. C.J. Geyer. Markov chain Monte Carlo maximum likelihood. In *Proceedings of the 23rd Symposium on the Interface*, page 156, New York, 1991. American Statistical Association.

8. P. Grassberger. Pruned-enriched Rosenbluth method: Simulation of θ polymers of chain length up to 1,000,000. *Phys. Rev. E.*, 56:3682–3693, 1997.

9. G. R. Grimmett and D. R. Stizaker. *Probability and Random Processes*. Oxford University Press, New York., 2001.

10. G. Güsoy, Y. Xu, A.L. Kenter, and J. Liang. Spatial confinement is a major determinant of the folding landscape of human chromosomes. *Nucleic Acid Res.*, 42:8223–8230, 2014.

11. W. K. Hastings. Monte Carlo sampling methods using markov chains and their applications. *Biometrika*, 57:97–109, 1970.

12. S. Henikoff and J. G. Henikoff. Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci.*, 89:10915–10919, 1992.

13. S. Karlin and S. F. Altschul. Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. *Proc. Natl. Acad. Sci.*, 87:2264–2268, 1990.

14. J. Liang, J. Zhang, and R. Chen. Statistical geometry of packing defects of lattice chain polymer from enumeration and sequential Monte Carlo method. *J. Chem. Phys.*, 117:3511–3521, 2002.

15. M. Lin, H.M. Lu, R. Chen, and J. Liang. Generating properly weighted ensemble of conformations of proteins from sparse or indirect distance constraints. *J Chem Phys*, 129(9):094101, 2008.

16. M. Lin, J. Zhang, H.M. Lu, R. Chen, and J. Liang. Constrained proper sampling of conformations of transition state ensemble of protein folding. *J Chem Phys*, 134(7):075103, Feb 2011.

17. P. Liò and N. Goldman. Models of molecular evolution and phylogeny. *Genome Res.*, 8:1233–1244, 1998.

18. J. S. Liu and R. Chen. Blind deconvolution via sequential imputations. *Journal of the American Statistical Association*, 90:567–576, 1995.

19. J. S. Liu and R. Chen. Sequential monte carlo methods for dynamic systems. *Journal of the American Statistical Association*, 93:1032–1044, 1998.

20. J. S. Liu, R. Chen, and T. Logvinenko. A theoretical framework for sequential importance sampling and resampling. In J.F.G. de Freitas A. Doucet and N. Gordon, editors, *Sequential Monte Carlo Methods in Practice*. Cambridge University Press, 2000.

21. J. S. Liu, R. Chen, and T. Logvinenko. A theoretical framework for sequential importance sampling and resampling. In A. Doucet, J. F. G. de Freitas, and N. J. Gordon, editors, *Sequential Monte Carlo in Practice*. Springer-Verlag, New York, 2001.

22. J. S. Liu, R. Chen, and W. H. Wong. Rejection control and importance sampling. *Journal of American Statistical Association*, 93:1022–1031, 1998.

23. H. Meirovitch. A new method for simulation of real chains: Scanning future steps. *J. Phys.A: Math. Gen.*, 15:L735–L741, 1982.

24. N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A.H. Teller, and E. Teller. Equations of state calculations by fast computing machines. *J. Chem. Phys*, 21:1087–1091, 1953.

25. C. P. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer-Verlag Inc., New York., 2004.

26. M. N. Rosenbluth and A. W. Rosenbluth. Monte Carlo calculation of the average extension of molecular chains. *J. Chem. Phys.*, 23:356–359, 1955.

27. G. Torrie and J. Valleau. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics*, 23(2):187–199, February 1977.

28. Y. Y. Tseng and J. Liang. Are residues in a protein folding nucleus evolutionarily conserved? *J. Mol. Biol.*, 335:869–880, 2004.

29. Y.Y. Tseng and J. Liang. Estimation of amino acid residue substitution rates at local spatial regions and application in protein function inference: A Bayesian Monte Carlo approach. *Mol. Biol. Evol.*, 23(2):421–436, Feb 2006.

30. S. Whelan and N. Goldman. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol. Biol. Evol.*, 18:691–699, 2001.

31. Z. Yang. Estimating the pattern of nucleotide substitution. *J. Mol. Evol.*, 39:105–111, 1994.

32. Z. Yang, R. Nielsen, and M. Hasegawa. Models of amino acid substitution and applications to mitochondrial protein evolution. *Mol. Biol. Evol.*, 15:1600–1611, 1998.

33. J. Zhang, R. Chen, C. Tang, and J. Liang. Origin of scaling behavior of protein packing density: A sequential monte carlo study of compact long chain polymers. *J. Chem. Phys.*, 118:6102–6109, 2003.

34. J. Zhang, M. Lin, R. Chen, and J. Liang. Discrete state model and accurate estimation of loop entropy of rna secondary structures. *J. Chem. Phys.*, 128(125107):1–10, 2008.

EXERCISES

3.1 In the Markov chain Monte Carlo method, the final stationary distribution reached after the chain convergency is the desired target contribution:

$$\int \pi(\boldsymbol{x})A(\boldsymbol{x},\boldsymbol{y})d\boldsymbol{x} = \pi(\boldsymbol{y}),$$

where $\boldsymbol{x}$ is the state variable, $A(\boldsymbol{x},\boldsymbol{y}) = T(\boldsymbol{x},\boldsymbol{y}) \cdot r(\boldsymbol{x},\boldsymbol{y})$ is the actual transition function, the product of the proposal function $T(\boldsymbol{x},\boldsymbol{y})$, and an acceptance-rejection rule $r(\boldsymbol{x},\boldsymbol{y})$. The proposal function $T(\boldsymbol{x},\boldsymbol{y})$ suggests a possible move from $\boldsymbol{x}$ to $\boldsymbol{y}$. The acceptance-rejection rule decides whether the proposed move to $\boldsymbol{y}$ will be accepted: Draw a random number u from the uniform distribution $\mathcal{U}[0,1]$. If $u \leq r(\boldsymbol{x},\boldsymbol{y})$, the move is accepted and $\boldsymbol{y}$ is taken as the new position. Otherwise stay with $\boldsymbol{x}$.

In the original Metropolis Monte Carlo method, the proposal function is symmetric: $T(\boldsymbol{x},\boldsymbol{y}) = T(\boldsymbol{y},\boldsymbol{x})$, and the acceptance-rejection rule is simply:

$$r(\boldsymbol{x},\boldsymbol{y}) = \min\{1, \pi(\boldsymbol{y})/\pi(\boldsymbol{x})\}$$

. Since the target distribution is the Boltzmann distribution $\pi(\boldsymbol{x}) \sim \exp(h(\boldsymbol{x}))$, where $h(\boldsymbol{x})$ is an energy function, the acceptance rule is often written as: $u \leq r(\boldsymbol{x},\boldsymbol{y}) = \exp(-[h(\boldsymbol{y}) - h(\boldsymbol{x})])$. This strategy will work, for example, if the proposal function gives equal probability $1/n(\boldsymbol{x})$ to each of the $n(\boldsymbol{x})$ conformations that can be reached from conformation $\boldsymbol{x}$:

$$T(\boldsymbol{x},\boldsymbol{y}) = 1/n(\boldsymbol{x}),$$

and if $n(\boldsymbol{x}) = n(\boldsymbol{y})$ for $\boldsymbol{x}$ and $\boldsymbol{y}$ that are connected by a move.

However, the number of possible moves for a conformation $\boldsymbol{x}$ frequently depends on the local geometry. For example, it is more difficult in protein simulation to move an amino acid residue that is buried in the interior than moving a residue located in a loop region. In other words, the number of allowed moves is different: $n(\boldsymbol{x}) \neq n(\boldsymbol{y})$, although each can be computed exactly.