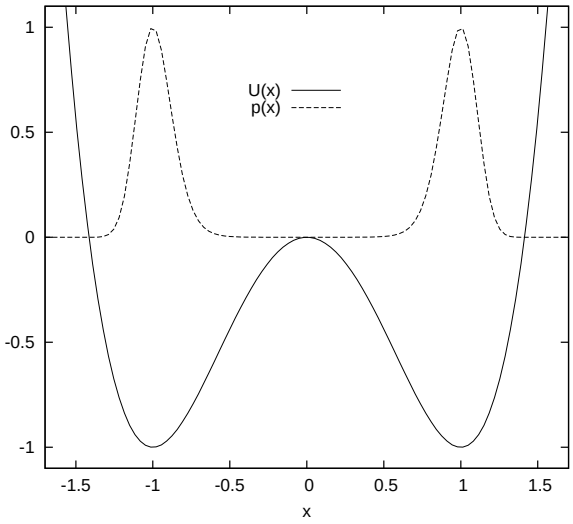
3

Applications of the Monte Carlo method

3.1 Quasi-ergodic sampling

- Detailed balance condition and ergodic transition matrix imply that random walk Monte-Carlo method correctly generates distribution of configurations.
- Says nothing about the *rate* of convergence, which depends on implementation (eigenvalue spectrum of transition matrix).
- Consider a 1-dimensional system with a quartic potential $U(x) = x^2(x^2 - 2)$.
- For $\beta = 10$, probability density $P(x) \sim e^{-\beta U(x)}$ is bimodal with small relative probability to be found near $x = 0$.



– Note that $P(1)/P(0) = e^{10}$.

- Consider the simple Metropolis Monte-Carlo scheme discussed previously with:

$$\mathsf{T}(x \to y) = \begin{cases} \frac{1}{2\Delta x} & \text{if } y \in [x - \Delta x, x + \Delta x] \\ 0 & \text{otherwise.} \end{cases}$$

– If maximum displacement Δx is small ($\Delta x \ll 1$), then if x is in the region near $x = 1$, the probability of proposing y near -1 is zero, and the proposal is very unlikely to propose a configuration y which is in a different mode from x.

– Random walk dynamics consists of long periods of x being localized in one of the modes, with only rare transitions between the modes (roughly every e^{10} steps). Overall, an equal amount of time must be spent in both of the modes.

– Rare transitions between modes leads to very slow convergence of distribution of states to $P(x)$.

– Easy to fix here since we know where the modes lie: increasing Δx will allow proposals between modes.

3.2 Umbrella sampling

- Origin of quasi-ergodic sampling problem is poor movement in random walk dynamics between different modes of high probability density.

- Sampling can be improved if one improves the transition probability between modes by either improving the proposal of trial moves or by modifying the acceptance criterion (i.e. sampling with a different importance function).

- **Umbrella sampling** is based on using a modified version of the Boltzmann distribution, typically specified by an additional potential energy term that encourages movement between modes.

- Consider the quartic system discussed in the last section. We define the *umbrella potential* $U_b(x) = kx^2$ and the importance function $\Pi(x) = e^{-\beta U(x)} e^{-\beta U_b(x)}$ and define the transition matrix so that Π is the limit distribution of the random walk.

– Any canonical average can be written as an average over Π

$$\langle A(x) \rangle = \int dx\, P(x) A(x) = \int dx\, \Pi(x) \left(A(x) \frac{P(x)}{\Pi(x)} \right)$$

provided that $\Pi(x) \neq 0$ at any point where $P(x) \neq 0$.

  - Generate Markov chain of states $\{x_1, \ldots, x_N\}$ according to $\Pi(x)$ so that an estimator of the average is

$$\langle A(x) \rangle = \frac{1}{N} \sum_{i=1}^{N} A(x_i) \frac{P(x_i)}{\Pi(x_i)} = \frac{1}{N} \sum_{i=1}^{N} A(x_i) w(x_i) = \frac{1}{N} \sum_{i=1}^{N} A(x_i) e^{\beta U_b(x_i)}.$$

  - Weight factor $w(x_i) = e^{\beta U_b(x_i)}$ accounts for bias introduced by the umbrella potential. In this case, it will assign greater weight to regions around the modes at $x = \pm 1$ since the biased random walk attaches less signficance to these regions.
  - The parameter k in the umbrella potential can be adjusted to minimize the statistical uncertainties.

- Disadvantage of umbrella approach: Must know a way in which to enhance movement between all modes of system in order to define an effective umbrella potential.

3.3 Simulated annealing and parallel tempering

3.3.1 High temperature sampling

- At high temperatures ($\beta \ll 1$), the equilibrium distribution $P_h(x)$ is only weakly bimodal.
  - Transition rate between modes depends exponentially on $\beta \Delta U$, where ΔU is the barrier height of the potential separating different modes.
  - If β is small so that $\beta \Delta U \ll 1$, then the system moves easily between modes.

- Can use high-temperature distribution $P_h(x)$ as an importance function $\Pi(x)$, resulting in weight function $w(x_i)$ given by

$$w(x_i) = \frac{P(x_i)}{\Pi(x_i)} = e^{-(\beta - \beta_h) U(x_i)} = e^{-\Delta \beta U(x_i)}.$$

  - If $\Delta \beta$ is large, points near barrier $x_i = 0$ receive little weight since $w(x_i) \ll 1$.
  - Advantage of this approach is that we don't need to know barrier locations since high average potential energy overcomes barriers.
  - Disadvantage: For high-dimensional systems, the number of accessible states is large (high entropy) for high temperatures. Many configurations sampled at high temperatures therefore receive little weight, leading to sampling inefficiency and large statistical uncertainties.

3.3.2 Extended state space approach: "Simulated Tempering", Marinari and Parisi, 1992

- Large temperature gaps in high temperature sampling approach lead to inefficient sampling due to a difference in density of states (entropy), while small temperature gaps are typically insufficient to enhance the passage between modes.

- Idea of extended state space approaches is to use a ladder of different temperatures (or other parameter) to allow the system to gradually move out of modes in an efficient manner.

- We augment the phase space point $\mathbf{r}^{(N)}$ with a parameter β_i from a set of m values $\{\beta_i\}$ and define a target limit distribution $\Pi(\mathbf{r}^{(N)}, \beta_i) = W_i e^{-\beta_i U(\mathbf{r}^{(N)})}$ on the extended state space $(\mathbf{r}^{(N)}, \beta_i)$, where W_i is an adjustable parameter.

- Monte-carlo procedure is standard, but with extended state space:

  1. Carry out sequence of a specified number of updates at fixed β_i using normal Metropolis scheme.
  2. Randomly and uniformly select a temperature index j, with corresponding parameter β_j, for a Monte-Carlo update. Accept change of parameter with probability $\mathsf{A}(i \to j)$, where

$$\mathsf{A}(i \to j) = \min\left(1, \frac{\Pi(\mathbf{r}^{(N)}, \beta_j)}{\Pi(\mathbf{r}^{(N)}, \beta_i)}\right) = \min\left(1, \frac{W_j}{W_i} e^{-(\beta_j - \beta_i)U(\mathbf{r}^{(N)})}\right).$$

  – Generate chain of states of extended phase space $\{(\mathbf{r}_1^{(N)}, i_1), \ldots, (\mathbf{r}_n^{(N)}, i_n)\}$. If target average is at temperature $\beta = \beta_1$, averages are given by estimator

$$\langle A(\mathbf{r}^{(N)})\rangle = \frac{1}{n}\sum_{k=1}^{n} A(\mathbf{r}_k^{(N)}) \frac{W_1}{W_{i_k}} e^{-(\beta_1 - \beta_{i_k})U(\mathbf{r}_k^{(N)})}.$$

- Drawbacks:

  – Must specify the parameter set $\{\beta_i\}$ properly to ensure the proper movement between modes.

  – Must know how to choose weights W_k for a given set of $\{\beta_k\}$. This can be done iteratively, but requires a fair amount of computational effort.

3.3.3 Parallel Tempering or Replica Exchange, C.J. Geyer, 1991

- Use an extended state space composed of replicas of the system to define a Markov chain $\mathbf{X} = (\mathbf{r}_1, \ldots \mathbf{r}_m)$, where each $\mathbf{r}_i$ is a complete configuration of the system.
- Design a transition matrix so that limiting distribution is

$$P(\mathbf{X}) = \Pi_1(\mathbf{r}_1) \ldots \Pi_m(\mathbf{r}_m)$$

  - The (statistically independent) individual components i of the extended state space vector can be assigned any weight Π_i. One choice is to use a Boltzmann distribution $\Pi_i(\mathbf{r}_i) = e^{-\beta_i U(\mathbf{r}_i)}$ with inverse temperature β_i.
- The Monte-Carlo process on the extended state space can be carried out as follows:
  1. Carry out a fixed number of updates on all replicas, each with a transition matrix K_i that has a limit distribution Π_i.
  2. Attempt a swap move, in which different components of the extended state space vector (replicas) are swapped.
     - For example, any pair of components, possibly adjacent to one another, can be selected from a set of all possible pairs with uniform probability. Suppose one picks components 2 and 3, so that the original configuration $\mathbf{X}_i$ and proposed configuration $\mathbf{Y}_i$ are

$$\mathbf{X}_i = \begin{pmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \mathbf{r}_3 \\ \vdots \\ \mathbf{r}_m \end{pmatrix} \qquad \mathbf{Y}_i = \begin{pmatrix} \mathbf{r}_1 \\ \mathbf{r}_3 \\ \mathbf{r}_2 \\ \vdots \\ \mathbf{r}_m \end{pmatrix}$$

     - The proposed configuration $\mathbf{Y}_i$ should be accepted with probability $\mathsf{A}(\mathbf{X}_i \to \mathbf{Y}_i)$ given by

$$\mathsf{A}(\mathbf{X}_i \to \mathbf{Y}_i) = \min\left(1, \frac{P(\mathbf{Y}_i)}{P(\mathbf{X}_i)}\right) = \min\left(1, \frac{\Pi_2(\mathbf{r}_3)\Pi_3(\mathbf{r}_2)}{\Pi_2(\mathbf{r}_2)\Pi_3(\mathbf{r}_3)}\right)$$

     - Note that no adjustable weight factors W_i are needed.
     - If $\Pi_i = e^{-\beta_i U(\mathbf{r}_i)}$, then

$$\frac{\Pi_2(\mathbf{r}_3)\Pi_3(\mathbf{r}_2)}{\Pi_2(\mathbf{r}_2)\Pi_3(\mathbf{r}_3)} = \frac{e^{-\beta_2 U(\mathbf{r}_3)}}{e^{-\beta_2 U(\mathbf{r}_2)}} \frac{e^{-\beta_3 U(\mathbf{r}_2)}}{e^{-\beta_3 U(\mathbf{r}_3)}} = e^{-(\beta_2-\beta_3)U(\mathbf{r}_3)} e^{(\beta_2-\beta_3)U(\mathbf{r}_2)} = e^{\Delta\beta\Delta U},$$

       where $\Delta\beta = \beta_3 - \beta_2$ and $\Delta U = U(\mathbf{r}_3) - U(\mathbf{r}_2)$.
- Each component of $\mathbf{X}_i$ in Markov chain of extended states $\{\mathbf{X}^{(1)}, \ldots \mathbf{X}^{(N)}\}$ is distributed with weight Π_i

- Averages at any of the temperatures are therefore readily computed using

$$\langle A \rangle_{\beta_i} = \frac{1}{N} \sum_{k=1}^{N} A(\mathbf{X}_i^{(k)}),$$

  where $\mathbf{X}_i^{(k)}$ is the ith component of the extended configuration $\mathbf{X}^{(k)}$ (the kth configuration in the Markov chain).

- Advantages: Quasi-ergodic sampling is mitigated by using a range of parameters such as β_i. The parameters should be defined such that their extreme values (such as highest temperature) should demonstrate no trapping in single modes.

- Disadvantages: There must be some overlap in adjacent densities Π_i and Π_{i+1} if a swap move is to be accepted with significant probability. Ideally, the parameters β_i should be chosen so that any given labelled configuration spends an equal amount of time at all parameter values.

  - Sometimes requires many replicas, which means that it takes many Monte-Carlo exchange moves for a given configuration to cycle through the parameters.

- One of the major advantages of the replica exchange method is the ease with which one can parallelize the algorithm.

  - Normal single-chain algorithm cannot be parallelized efficiently because of sequential nature of random walk procedure.
  - Can parallelize the generation of many trial configurations and use an asymmetric proposal procedure to select one preferentially.
  - Can parallelize computation of energy, if this is a rate-determining step.

- The exchange frequency between replicas can be optimized to suit the computer architecture.

- Each processor can deal with a single replica, or sub-sets of replicas.