Advanced topics in Markov-chain Monte Carlo

Lecture 7:
Meta algorithms (a practical approach)
Part 1/2: Simulated annealing, following Hajek (1988)

Werner Krauth

ICFP -Master Course Ecole Normale Supérieure, Paris, France

9 March 2022

References

- D. A. Levin, Y. Peres, E. L. Wilmer, **Markov Chains and Mixing Times**, (American Mathematical Society, 2008),

Characteristic times in MCMC 1/2

- Correlation time: Time to move from one i ($\sim \pi$) to an independent j ($\sim \pi$).
- Mixing time: Time to reach a $j \sim \pi$ (to some precision) starting from $i \sim \pi^{\{0\}}$ with worst $\pi^{\{0\}}$.
- Cover time: Time to have seen all samples, starting from the worst initial sample x: $t_{\text{cov}} = \max_{x \in \Omega} \mathbb{E}\left[\tau_{\text{cov}}(x)\right]$ (with $\tau_{\text{cov}}(x)$ the time to have seen all $i \in \Omega$).

Characteristic times in MCMC 2/2

Example (SSEP of N hard spheres on path graph $\mathcal{P}_{2N}$):

- Correlation time: $\propto N^3$.
- Mixing time: $\propto N^3 \log N$.
- Cover time: N^N.

Consequences:

1. Difficult to know normalization of π:
   - ... What is $Z = \sum_x \pi_x$? (Thermodynamic integration)
2. Difficult to know Ω:
   - ... What is $\min_x \pi_x$?
   - ... What is $\max_x \pi_x$? (Simulated annealing)
   - ... What is conductance?
3. Difficult to explore Ω:
   - Is $\Omega = \emptyset$?
   - Have we seen all of Ω? (Multicanonical MC)

Simulated annealing 1/5

Simulated annealing: MCMC Optimization algorithm

- Start at very small values of α,
- At each step, slowly increase α by a tiny amount.

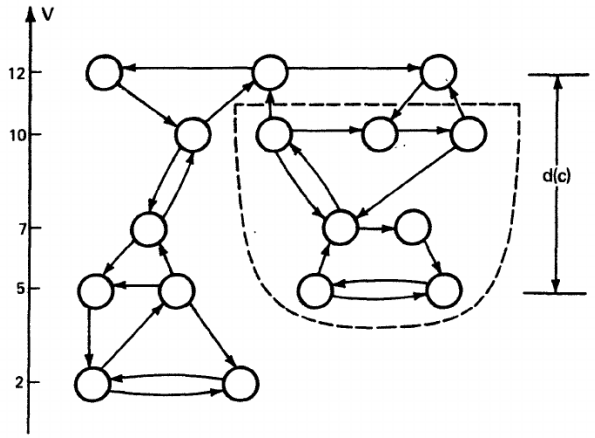
In the V-shaped probability distribution (with $\pi(n) = 0^+$), switch to temperature language.

- $U_i = -\log(\pi_i)$
- Set up temperature schedule $T_k \to 0$ for $k \to \infty$
- Accept / reject move ΔU with Metropolis filter $\min[1, \exp(-\Delta U / T_k)]$

Simulated annealing 2/5

Simulated annealing: MCMC optimization algorithm

- Theorem (Hajek 1988): need $\sum_k \exp\left(-d^*/T_k\right) = \infty$ for sure convergence to lowest-energy configuration.
- Corrolary (Hajek 1988): If $T_k = c/\log\left(k+1\right)$, then need $c \geq d^*$
- Easy to check in V-shaped distribution on the path graph.



Simulated annealing 3/5

```
weight = [0.0]
potential = [1000.0]
for i in range(1, L + 1):
    weight.append(abs( (L + 1.0) / 2.0 - i))
    potential.append(-math.log(weight[-1]))
weight[-1] = 0.0
potential[-1] = 1000.0
```

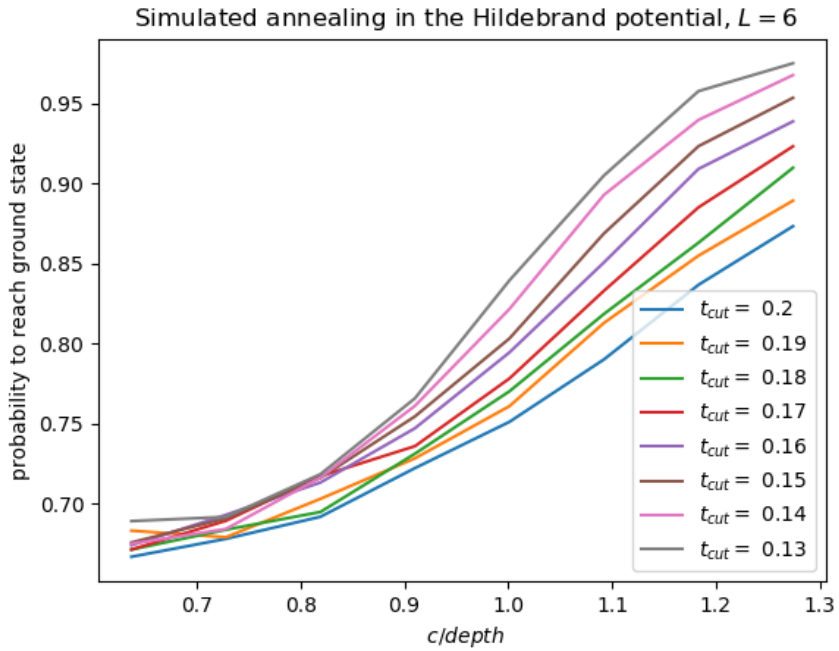
- Weights and potentials (energies).
- Identifying the minimal

Simulated annealing 4/5

```
for cutoff in [0.20,0.19,0.18,0.17,0.16,0.15,0.14,0.13]:
    xvalue = []
    yvalue = []
    for c in [0.70, 0.8,0.9,1.0,1.1,1.2,1.3,1.4]:
        SuccessRate = 0.0
        for attempt in range(Nsample):
            x = L // 2
            t = 1
            while True:
                t += 1
                Temp  = c / math.log(1.0 + t)
                if Temp < cutoff: break
```

- Logarithmic annealing schedule.
- Scale set by maximal barrier.
- Low-temperature cutoff.

Simulated annealing 5/5



- Probability to reach groundstate for the $L = 6$ Hildebrand model.
- NB: x-axis—annealing scale / potential scale.
- TODO: Follow $\pi^{\{t\}}$ with the time-dependent transition matrix (TD).