Survey Paper

Take a bunch of local random decisions that converge to a global state that adheres to a required distribution.

#P → class of hard counting problems.
MC sampling: Allows us to approximate answers to these sort of problems.

Consider a simple MC for Indep. Set.

i) choose $v \in_R V$
ii) If $v \in I_t$ then move to state $I_t - \{v\}$.
iii) Else if $v \notin I_t$ and $\{v\} \cup I_t$ is a valid indep. set then move to $I_t \cup \{v\}$
iv) Else dont move.

Looking at the transition prob -

Hamming Dist

$$
Pr[I, I'] = \begin{cases} 1/n, & \text{if } d(I, I') = 1 \\ 0, & \text{if } d(I, I') > 1 \\ 1 - \sum_{J \neq I} Pr[I, J], & I = I' \end{cases}
$$

Definition: MC is ergodic:

i) irreducible: $\forall x, y \ \exists t$ s.t. $p(x, y) > 0$.
(at some point I can reach anywhere from anywhere with non-zero probability.)

ii) aperiodic: $\forall x, y \ \gcd(\{t \mid p^t(x, y) > 0\} = 1$.
(there is no bipartition, a bipartition would cause one side of the cut to be non zero at odd timesteps and

the other side at even time steps.).

Notice that the indep set chain is both. Particularly, the self loop makes sure that bipartition is impossible.

This can work on any chain - add self loops with some const prob and it will aperiodic (however there may be a small, constant cost to the runtime).

Lazy chain: $\frac{1}{2}$ prob of self loop at each state.

Main Lemma: Any finite, ergodic MC converges to a unique stationary distribution π, i.e. $\forall x, y$ as $t \to \infty$ $p^t(x,y) = \pi(y)$.

Detailed Balance:
For an ergodic MC with trans prob P, if $\pi' : \Omega \to [0,1]$ is any function s.t.

$$\pi'(x) \cdot P(x,y) = \pi'(y) \cdot P(y,x) \text{ s.t.}$$

$\sum_{x \in \Omega} \pi'(x) = 1$ then π' is the unique stationary distribution.

Chains satisfying detail balance: time reversible.

Notice that for the indep set chain, $p(I, I') = p(I', I) = \frac{1}{n}$. for $I \neq I'$. $\Rightarrow$ detailed balance is satisfied as.

$$\frac{1}{n} \cdot \frac{1}{|\Omega|} = \frac{1}{n} \cdot \frac{1}{|\Omega|} . \Rightarrow \pi'(x) = \frac{1}{|\Omega|}$$

uniform distribution

Natural Questions -

i) can I choose a more complicated distribution.

ii) How long does it take to converge.

<u>Metropolis Algorithm</u>: Answer to question 1.

Starting at x repeat:

i) Pick a neighbour y of x with prob $\frac{1}{2\Delta}$.

ii) Move to y with prob $\min\left(1, \frac{\pi(y)}{\pi(x)}\right)$.

iii) With remaining prob stay at x.

Claim: Detail Balance is satisfied.

Verify: $\quad LHS = \frac{1}{2\Delta} \min\left(1, \frac{\pi(y)}{\pi(x)}\right) \cdot \pi(x)$

$$RHS = \frac{1}{2\Delta} \min\left(1, \frac{\pi(x)}{\pi(y)}\right) \cdot \pi(y)$$

Now if $\pi(x) > \pi(y)$

then $LHS = \frac{1}{2\Delta} \pi(y)$, $\quad RHS = \frac{1}{2\Delta} \cdot 1 \cdot \pi(y)$

Hence equal

Similarly if $\pi(x) < \pi(y)$ also.

Going back to the IS example, suppose we want to sample with prob $\frac{\lambda^{|I|}}{Z}$ where $Z = \sum_{\forall I} \lambda^{|I|}$ is the normalizing constant.

consider I and $I' = I \cup \{v\}$ in the MC.

Here $|I'| = |I| + 1$

$\Rightarrow \pi(I') = \lambda \pi(I)$

$$\Rightarrow P(I, I') = \frac{1}{2n} \min\left(\frac{\lambda \pi(I)}{\pi(I)}, 1\right) = \frac{1}{2n} \min(1, \lambda)$$

But $P(I', I) = \frac{1}{2n} \min\left(1, \frac{1}{\lambda}\right)$.

Notice that Z is gone. (We don't know how to count it any way, thats a #P problem.

Again, verifying detailed balance

$$LHS = \pi(I) \cdot \frac{1}{2n} \min(1, \lambda)$$

$$RHS = \lambda\pi(I) \cdot \frac{1}{2n} \min\left(1, \frac{1}{\lambda}\right)$$

If $\lambda < 1$, $LHS = \frac{\lambda\pi(I)}{2n}$, $RHS = \frac{\lambda\pi(I)}{2n} \cdot 1$.

$\lambda \geq 1$, $LHS = \frac{\pi(I)}{2n}$, $RHS = \frac{\lambda\pi(I)}{2n\lambda} = \frac{\pi(I)}{2n}$

In both cases $LHS = RHS$.

<u>Mixing Time</u>

Time taken for an MC to converge to its stationary distribution

Defn: Total Variational Distance at time t

$$\| P^t, \pi \|_{tv} = \max_{x\in\Omega} \frac{1}{2} \sum_{y\in\Omega} | P^t(x,y) - \pi(y) |$$

Choose the worst source and find L1 norm. and divide by 2.

Defn: For $\varepsilon > 0$, mixing time $\tau(\varepsilon)$ is

$$\tau(\varepsilon) = \min \{ t \mid \| P^t(x,y), \pi \|_{tv} \leq \varepsilon, \forall t' \geq t \}$$

Rapidly Mixing: If $\tau(\varepsilon)$ is bounded by poly$(n, \log(\varepsilon))$ where n is the size of each configuration.
(for ex - #nodes in the graph for the IS chain).

For $\lambda_0, \lambda_1, \cdots, \lambda_{|\Omega|-1}$ eigen values of the transition matrix s.t. $1 = \lambda_0 > |\lambda_1| \geq |\lambda_i| \; \forall i$ then-

Defn: Spectral Gap $\lambda_0 - |\lambda_1| = 1 - \lambda_1$.

Theorem: Let $\pi_* = \min\limits_{x \in \Omega} \pi(x)$. For all $\varepsilon > 0$,

a) $\tau(\varepsilon) \leq \dfrac{1}{1 - |\lambda_1|} \log\left(\dfrac{1}{\pi_* \varepsilon}\right)$

b) $\tau(\varepsilon) \geq \dfrac{|\lambda_1|}{2(1 - |\lambda_1|)} \log\left(\dfrac{1}{2\varepsilon}\right)$

Unfortunately this neat trick isn't useful when the transition prob matrix is unavailable. i.e. when the state space is exponentially large.

Coupling

Defn: A MC on $\Omega \times \Omega$ defining $(X_t, Y_t)_{t=0}^{\infty}$ s.t.
i) X_t and Y_t are copies of M
ii) If $X_t = Y_t$ then $X_{t+1} = Y_{t+1}$.

i.e. i) Each when viewed individually simulates the original chain.
ii) Once they both agree, they agree from that time forward.

Defn: For initial states x, y

$$T^{x,y} = \min\{t \mid X_t = Y_t \text{ for } X_0 = x, Y_0 = y\}$$

i.e. the first time they agree.

Defn: Coupling Time $T = \max_{x,y} E[T^{x,y}]$

Thm: $\tau(\epsilon) \le \lceil T e \ln \epsilon^{-1} \rceil$

Toy Example: Sampling a point from a boolean hypercube. The natural flip a bit UAR won't work since its bipartite (odd and even # bits set). So a modified chain -

MCcube

starting at $X_0 = (0, 0, \cdots, 0)$ do

i) Pick $(i, b) \in \{1, \cdots, n\} \times \{0, 1\}$

ii) $X^{t+1} = X^t$ with $X^t[i] = b$.

This gives us transition prob $P(X,Y) = \begin{cases} \frac{1}{2n}, & d(X,Y) = 1 \\ \frac{1}{2}, & X = Y \\ 0, & o/w \end{cases}$

Now if you want to couple, start from two vertices X_0 and Y_0 update these with the same (i, b).

$\Rightarrow$ If we have updated each bit at least once, by pigeonhole principle they must have collided.

Apparently this converges in $\Theta(n \log n)$. due to coupon collector

$\Rightarrow$ Coupling Time : $n \log n$

$\Rightarrow$ Mixing Time $\tau(\varepsilon) = n \dfrac{\ln n}{\ln 2} e \ln \varepsilon^{-1}$

$$= O(n \ln(n\varepsilon^{-1}))$$

Coupling proofs may get more complicated if the distance b/w configurations increases. For the hypercube this distance is non decreasing (because the update is the same). When it can increase, the argument doesn't work directly, but if in expectation it decreases, we can use some theory of random walks to show an $n \log n$ coupling time.

Path Coupling

Finding expected change in distance is not straightforward.

Bubley and Dyer showed that it is sufficient to consider only pairs of configurations that are close.

i.e. consider a small set $U \subseteq \Omega \times \Omega$ that are close in some distance metric, ϕ. For ex- pairs of configs with hamming distance 1. Now suppose expected change in distance is decreasing for all pairs in U.

To reason about some arbitrary $X, Y \in \Omega$. Let $z_0 = X \cdots z_r = Y$ define a shortest path from X to Y of length r, where $(z_i, z_{i+1}) \in U$.

$\Rightarrow \phi(X,Y) = \sum_{i=0}^{r-1} \phi(z_i, z_{i+1})$. Each of these individual decrease in expectation. Then by linearity $\phi(X,Y)$ also decreases in expectation.

Formally: Let Φ be an integer valued metric defined on $\Omega \times \Omega$ which takes values in $\{0, \dots, B\}$. Let U be a subset of $\Omega \times \Omega$ such that for s.t for all (x_t, y_t) in $\Omega \times \Omega$ there exists a path $x = z_0, \dots, y = z_r$ s.t $(z_i, z_{i+1}) \in U$. and

$$\sum_{i=0}^{r-1} \phi(z_i, z_{i+1}) = \phi(x_t, y_t)$$

Let M be an MC on Ω with transtion matrix P. consider a random function $f : \Omega \to \Omega$, such that $\Pr[f(x) = y] = P(x, y)$.

Define a coupling of the MC by $(x_t, y_t) \to (x_{t+1}, y_{t+1}) = (f(x_t), f(y_t))$

i. If there exists $\beta < 1$ s.t. $E[\phi(x_{t+1}, y_{t+1})] \leq \beta\, \phi(x_t, y_t)$ for all $(x_t, y_t) \in U$ then the mixing time satisfies –

$$\tau(\varepsilon) \leq \frac{\ln(B\varepsilon^{-1})}{1-\beta} \quad (B \text{ is max metric})$$

ii. If $\beta = 1$: If $\exists\, \alpha > 0$ s.t $\Pr[\phi(x_{t+1}, y_{t+1}) \neq \phi(x_t, y_t)] \geq \alpha$. For all x_t, y_t then

$$\tau(\varepsilon) \leq \left\lceil \frac{eB^2}{\alpha} \right\rceil \lceil \ln \varepsilon^{-1} \rceil$$

Glauber Dynamics

1. with prob $= 1/2$ do nothing.
2. $(v, c) \in_R V \times \{1, \dots, k\}$

3. If v recolored with c is valid recolor else do nothing.

Transition matrix is symmetric $\Rightarrow$ converges to uniform distribution over all valid k-colorings.

Lets look at coupling here. Let X_0, Y_0 be two arbit, valid k-colorings. Lets update both with the same update (v, c).

<u>Thm</u>: $T(\varepsilon) = O(n \ln(n\varepsilon^{-1}))$ for an n-vertex graph with $k \geq 3\Delta + 1$.

<u>Proof</u>: Let x_0, y_0 be starting configurations. At each t, choose (v, c) UAR and update both x^t an y^t as defined.

Let $\phi : \Omega \times \Omega \to \mathbb{Z}$ be the hamming metric. i.e. the min no. of colors to flip to reach one config from another. Here notice that the max possible shortest distance $B = 2n$ (not sure - won't $n+1$ be enough?).

Let U be the set of pairs of colorings at $\phi = 1$. Now we are interested in change in ϕ (r, s), $(r, s) \in U$.

Suppose w is the vertex that differs in r and s,

i) If $w = v$, neighborhood is same, hence either distance decreases (if both accept the c) or remains same. There are atleast $k - \Delta$ colors that will be accepted.

$$\Rightarrow \mathbb{E}[\Delta\phi] \leq -1 \times \frac{(k-\Delta)}{kn} + 0 \times \left(1 - \left(\frac{k-\Delta}{kn}\right)\right)$$

$$\leq -\frac{(k-\Delta)}{kn}.$$

ii) $(\omega, v) \in E$: If they both accept or reject the distance is unchanged. However if $c = \chi(v)$ for either, the distance will increase by 1.

$$\Rightarrow \mathbb{E}(\Delta\phi) \leq \frac{2}{kn} \quad \text{for fixed } \omega.$$

$\Rightarrow$ for Δ possible ωs, $\mathbb{E}(\Delta\phi) \leq \frac{\Delta 2}{kn}$

iii) $\omega \neq v$ and $(\omega, v) \notin E$: Here the change in dist is always 0 since both will either accept or reject.

$$\Rightarrow \mathbb{E}[\Delta\phi] \leq \frac{1}{kn}(-k + \Delta + 2\Delta) = \frac{3\Delta - k}{kn}$$

$\Rightarrow$ if $k > 3\Delta$, $\mathbb{E}(\Delta\phi) < 1$

$$\Rightarrow \tau(\varepsilon) \leq \frac{\ln 2n\varepsilon^{-1}}{1 - \frac{3\Delta - k}{kn}}$$

For tight bound let $k = 3\Delta + 1$

$$\Rightarrow \tau(\varepsilon) \leq \frac{\ln n\varepsilon^{-1}}{1 - \frac{1}{kn}}$$

$$= O(n \ln(n\varepsilon^{-1}))$$

Extensions

→ choosing better coupling.

Glauber Dynamics can actually proven to converge with the above time even if $k > 2\Delta$. This achieved by looking carefully at case ii.

→ Changing the Markov chain.

Possibly add extra moves or update more than one node at a time. Anything that makes analyzing these distances more efficient. In fact, it is proved that rapid mixing of some modified chain implies rapid mixing of the original Glauber dynamics.

→ Macromoves : Add a "warm up" or "burn in" period that randomizes the configurations before each step.

Vigoda and Hayes have extended this concept to work during the coupling phase itself.

Flows

Intuition : slow mixing time is caused by some sort of bottleneck ⇒ exponentially small cut.

Min cut

Conductance of a chain –

For $S \subset \Omega$, let

$$\phi_S = \sum_{\substack{x \in S, \\ y \notin S}} \frac{Q(x,y)}{\pi(S)}$$

where $Q(x,y) = \pi(x) \cdot P(x,y)$. This can be thought of as the capacity of edge (x,y). And $\pi(S) = \sum_{x \in S} \pi(x)$ is the weight of the cut set. (weight of edges crossing the cut). (I don't see how).

Conductance $\Phi = \min_{S \mid \pi(S) \le \frac{1}{2}} \phi_S$

Intuitively - low conductance $\Rightarrow$ there is a bad cut.

since Φ_S can be thought of as a cut divided by $\pi(S)$. This is sort of a min cut.

Thm: For MC with conductance Φ and eig val gap $Gap(P) = 1 - |\lambda_1|$,

$$\frac{\Phi^2}{2} \leq Gap(P) \leq 2\Phi.$$

We know $Gap(P)$ relates to $\tau(\varepsilon)$ as-

$$\tau(\varepsilon) \leq \frac{1}{Gap(P)} \log\left(\frac{1}{\underbrace{\pi^*}_{\min_x \pi(x)}\varepsilon}\right)$$

$\Rightarrow$ Small Φ $\Rightarrow$ small Gap $\Rightarrow$ Larger $\tau(\varepsilon)$.

<u>Max Flow</u>

Consider $G(\Omega, E =$ non zero transitions$)$. Now min cut corresponds to a max flow

$\rightarrow$ for each pair of x, y define canonical path $\gamma_{x,y}$

$\Gamma = \{\gamma_{xy} \mid x, y \in \Omega, x \neq y\}$. (some arbit path)

$$\Rightarrow \text{congestion} \quad \rho(\Gamma) = \max_e \frac{1}{Q(e)} \sum_{\underbrace{\gamma_{x,y} \ni e}_{\text{some edge in the path.}}} \pi(x)\pi(y).$$

i.e. an edge appearing in more paths is more congested.

Think of each path from x to y as carrying $\pi(x)\pi(y)$ units of flow. congestion ρ measures the max ration of load routed through edge e divided by it capacity $Q(e)$.

Let $\bar{\rho} = \min_{\Gamma} \rho(\Gamma) . l(\Gamma)$ where l is the max length of any path in Γ.

Thm: For an ergodic, reversible MC with stationary π and self loop prob $P(y,y) \geq \frac{1}{2}$ $\forall y \in \Omega$

$$\tau_x(\varepsilon) \leq \bar{\rho}\left(\log \pi(x)^{-1} + \log \varepsilon^{-1} \right)$$

consider the hypercube chain. Choose Γ that makes paths by moving left to right and flipping bits as they mismatch.

$\rho(\Gamma)$: Consider an edge e.
Now $\sum_{e \in \gamma_{xy}} \pi(x)\pi(y)$. Now $\pi(x) = \pi(y) \Rightarrow$ just have to count the #path containing (u,v).

If u, v differ in the ith bit, first $i-1$ bits of u agree with the end of the path, y.

similarly last $n-i$ bits must agree with x, the start of the path.

$\Rightarrow x = (x_1 \cdots x_{i-1}, u_i, v_i \cdots v_n)$
$\quad y = (v_1 \cdots v_i, y_{i+1}, \cdots y_n)$

$\rightarrow$ 2^{n-1} ways to assign $x_1 \cdots x_{i-1}, y_{i+1} \cdots y_n$

$$\Rightarrow \sum_{e \in \Gamma_{xy}} \pi(x)\pi(y) = \frac{1}{2^n} \cdot \frac{1}{2^n} \cdot 2^{n-1} = 2^{-(n+1)}$$

Also, $Q(e) = \underset{\pi(x)}{\underset{\uparrow}{\frac{1}{2^n}}} \cdot \underset{P(x,y)}{\underset{\uparrow}{\frac{1}{2n}}}$

$$\Rightarrow \rho(\Gamma) = n 2^{n+1} \cdot 2^{-(n+1)} = n$$

$$l(\Gamma) \le n$$

$$\Rightarrow \tau(\varepsilon) \le n(\log 2^n + \log \varepsilon^{-1})$$

Extensions to flow-based analysis

Suppose we want to sample from the set of matchings in a graph, s.t.

$$\pi(M) = \frac{\lambda^{|M|}}{Z}$$ where Z is the usual normalizing const.

Consider an MC that does -

i) Choose an edge $e = (u,v)$ uniformly.

ii) If $e \in M$ remove it with prob $\min(1, \lambda^{-1})$

iii) If $e \notin M$ and both u and v are unmatched add it with prob $\min(1, \lambda)$.

iv) If exactly one of u and v are matched remove it and add edge (u,v).

v) Else do nothing.

This is sort of a Metropolis chain.

If X and Y differ by one edge -
and $|X| + 1 = |Y|$,

$$\pi(X) \cdot P(X,Y) = \frac{\mu^{|X|}}{Z} \cdot \frac{1}{m} \cdot \min(1, \mu)$$

$$\pi(Y) \cdot P(Y,X) = \frac{\mu^{|X|+1}}{Z} \cdot \frac{1}{m} \min(1, \mu^{-1})$$

If $\mu \leq 1$ (and similarly for $\mu > 1$).

$$LHS = \frac{\mu^{|X|}}{Z} \cdot \frac{1}{m} \cdot \mu = \frac{\mu^{|X|+1}}{Zm} = RHS.$$

If X and Y differ by 2 edges.
and $|X| = |Y|$

$$\pi(X) \cdot P(X,Y) = \frac{\mu^{|X|}}{Z} \cdot \frac{1}{m}$$

$$\pi(Y) \cdot P(Y,X) = \frac{\mu^{|Y|}}{Z} \cdot \frac{1}{m}$$

which are equal.

Unlike the hypercube example, we can't just count paths.

Let x, y be any two matchings in G. Consider $x \oplus y$ to be a graph with same vertex set but XOR of edge sets. This must be a collection of alternating cycles and paths.

To define a path from x to y on the chain we take the first component of $x \oplus y$.
Here a "step" involves removing an edge from x and (if exists) adding the alternate edge to X. Repeat until all original edges from X

in the component is removed.

Now consider an MC transition (u,v). How many paths from x to y pass through (u,v).

simultaneously consider a path from y to x. Now $y \oplus x$ is same as $x \oplus y$.

Now you can do some interesting stuff: explored in another paper to analyze this.

there is also a variant known as balanced flows.

Auxilary Methods -

Comparison: modify MC without significantly changing the mixing time.

Let $\tilde{P}$ and P be reversible MCs on the same state space Ω with the same stat. distr π.

If $\tau_{\tilde{P}}(\varepsilon)$ is known/bounded and we want $\tau_P(\varepsilon)$ we use -

For each x, y with $\tilde{P}(x,y) > 0$, define γ_{xy} with $P(x_i, x_{i+1}) > 0$. Let $|\gamma_{xy}|$ be length of this path.

Let $\Gamma(z,w) = \{ (x,y) \mid (z,w) \in \gamma_{xy} \}$ the set of paths that use (z,w) transition.

$$\text{Define } A = \max_{(z,w) \in E(P)} \left\{ \frac{1}{\pi(z) \cdot P(z,w)} \cdot \sum_{(x,y) \in \Gamma(z,w)} |\gamma_{xy}| \, \pi(x) \cdot \tilde{P}(x,y) \right\}$$

Thm: $\mathrm{Gap}(P) \geq \frac{1}{A} \mathrm{Gap}(\tilde{P})$

Issue: same state space. very restrictive.

Idea: <u>Decomposition</u>

Break down the chain into multiple pieces and relate overall runtime to those of the smaller chains.

Formally break $\Omega = \Omega_1, \ldots, \Omega_m$ disjoint subsets.

Transitions are - $P_i(x,y) = p(x,y)$ if $x, y \in \Omega_i$

$p_i(x,x) = 1 - \sum_{y \notin \Omega_i} p(x,y)$.

$$\pi_i(A \subseteq \Omega) = \frac{\pi(A \cap \Omega_i)}{\pi(\Omega_i)}$$

Now transitions b/w partitions are determined by—

$$\bar{P}(i,j) = \frac{1}{\pi(\Omega_i)} \sum_{\substack{x \in \Omega_i \\ y \in \Omega_j}} \pi(x) P(x,y).$$

<u>Thm</u>: $\mathrm{Gap}(P) \geq \frac{1}{2} \cdot \mathrm{Gap}(\bar{P}) * \min_{i=1,\ldots,m} \mathrm{Gap}(P_i)$

Commonly Studied Models

<u>1. Independent Sets</u>: G be a graph and Ω be the

set of independent sets in G. Think of each indep set as a map from V to $\{0,1\}$ and no two neighbours can both be colored 1.

Let x_0, x_1 be the "desirability" of having a vertex in or out of an IS. Define weight -

$$\omega(I) = \prod_{v \in V} x_{f(v)}$$

This means $\pi(I) = \dfrac{w(I)}{\sum_{I' \in \Omega} w(I')}$

2. Colorings

$f: V \to \{1, \dots, k\}$

Ω : set of proper k-colorings.

Let $x_0 \dots x_k$ be weights associated with each colors.

$$w(C) = \prod_{v \in V} x_{f(v)}$$

3. Matchings

$f: E \to \{0,1\}$

Again x_0, x_1 are weights.

$$\omega(M) = \prod_{e \in E} x_{f(e)}$$

if $x_0 = 1$, $x_1 = \mu$ for integer μ, $w(M)$ weights matchings as if G was a multigraph with μ parallel edges instead of each tree edge.

4. Pairwise influence models

For any n vertex $G(V,E)$, Let $\Omega = \{1, \cdots, q\}^n$ where
$f : V \rightarrow \{1, \cdots, q\}$

Define a symmetric set of weights $\{X_{ij} = X_{ji}\}$ for each pair $i, j \in \{1 \cdots q\}$ weight each configuration by

$$w(\sigma) = \prod_{u,v \in E} X_{f(u) f(v)}$$

Thus we can get ~~q~~-colorings that favour some edges.
For ex if $X_{ij} = 1$ for all $i \neq j$ and $X_{ij} = 0$ for $i = j$ gives uniform distribution on proper k-coloring.

Unifying Framework for these problems.

In statistical physics, models represent simple physical systems. Systems favor configurations that minimize energy.

The energy function on the space of configuration is determined by a Hamiltonian $H(\sigma)$.

$$w(\sigma) = \exp\left(-\beta H(\sigma)\right) \quad \text{where } \beta = \frac{1}{T}$$

For low values of β (high T) the diff b/w energies of configurations is dampered.
At high β (low T) the diff are magnified.

Likelihood of each configuration: $\pi(\sigma) = \frac{w(\sigma)}{Z}$

$Z = \sum_{\tau} w(\tau)$ known as partition function.

This overall distr is called Gibbs/Boltzmann distr.

Let state space be the set of indep sets of a graph.

$$H(I) = -\sum_{v \in V} \delta_{v \in I} = -|I|. \text{ where } \delta = \begin{cases} 1 & \text{if } v \in I \\ 0, & o/w \end{cases}$$

Now the prob distr -

$$\pi(I) = \frac{e^{-\beta H(I)}}{Z}$$

If $e^{\beta} = \lambda$, we get $\frac{\lambda^{|I|}}{Z}$. This is called the hard-core gas model.

Another model is the Ising model. Given a Graph with n vertices, look at the 2^n ways of assigning spin ± 1.

$$H(\sigma) = -\sum_{(u,v) \in E} \sigma_u \sigma_v = |D(\sigma)| - |A(\sigma)|$$

where D is the set of edges where $\sigma_u \neq \sigma_v$
A $\sigma_u = \sigma_v$

$$\text{Then } \pi(\sigma) = \frac{1}{Z} \exp\left(-\beta(|D(\sigma)| - |A(\sigma)|\right)$$

$$= \frac{1}{Z} \exp\left(-\beta(|E| - 2|A(\sigma)|\right)$$

$$= \frac{\exp(\beta|E|) \cdot \exp(-2\beta|A(\sigma)|)}{\sum_{\tau \in \{-1,1\}^n} \exp(\beta|E|) \cdot \exp(-2\beta|A(\sigma)|)}$$

$$= \frac{\exp(-2\beta|A(\sigma)|)}{\sum_{\tau} \exp(-2\beta|A(\tau)|)}$$

This is a special case of pairwise influence where $q = 2$, $X_{00} = X_{11} = \exp(-2\beta)$ and $X_{01} = X_{10} = 1$.

For larger q, the pairwise influence model specializes to the potts model.

Some Terms "translated"

Physics	CS
monomer-dimer coverings	Matchings
dimer coverings	perfect Matchings
Hard-core gas model	Independent set
spin	bit
Ground states	Highest prob configurations
Ground state in Potts	vertex coloring
Partition function	Normalizing constant Z.
connectivity	degree
activity / fugacity	vertex weights
interaction	edge weights
ferromagnetism	Positive correlation
anti-ferromagnetism	Negative correlation
mean field	K_n
Bethe Lattice	complete regular tree
polynomial mixing	
rapid mixing	$O(n\log n)$ mixing.