4.1 Many Random Vars I:

$$P_N(\underline{x}) = P_N(x_1, \dots, x_N)$$

$$r(\underline{x}) := \frac{1}{N} \log \frac{1}{P_N(\underline{x})}$$

$$\mathbb{E}_{\underline{x}}\, r(\underline{x}) = h_N$$

How does $r(\underline{x})$ fluctuate?

Turns out, as $N \to \infty$

$$P(r(\underline{x}) \approx \rho) \doteq e^{-N I(\rho)}$$

$I(h) = 0$ is the single min

We have e^{Nh} sequences of max probability all of roughly equal probability "equipartition"

compared to $e^{N \log |\mathcal{X}|}$ total sequences

4.2 large deviations for indep. vars

symbol x

$$q_{\underline{s}}(x) = \frac{1}{N} \sum_i \mathbb{1}(x = s_i)$$ empirical dist

$$\underset{\underline{s} \sim p}{P}[q] = \binom{N}{qN} p^{qN} (1-p)^{N(1-q)}$$ ⟸ Binomial dist

binary case

$$= \exp\left[N\left(\mathcal{H}(q) + q \log p + (1-q) \log(1-p)\right)\right]$$

$$= \exp\left[N\left(\mathbb{E}_{x\sim q} \log \frac{p(x)}{q(x)}\right)\right] = \exp\left[-N D_{KL}(q \| p)\right]$$

General case: k symbols

$$= \binom{N}{q_1 N \cdots q_k N} p_1^{q_1 N} \cdots p_k^{q_k N}$$

$$= \exp\left[N\left(\mathcal{H}(q_1, \cdots, q_k) + \sum q_i \log p_i\right)\right] = \exp[-N D_{KL}(q \| p)]$$

Eg 4.4 Atmosphere

$$E = \sum_{i=1}^{N} z_i \Rightarrow Z = \sum_{\{z_i\}} \exp[-\beta E(z_i)] = \prod_i \sum_z e^{-\beta z}$$

$$= \left(\frac{1}{1-e^{-\beta}}\right)^N$$

density profile, just like for eigs $\rightarrow$ $p(z) = \frac{1}{N} \sum_i \mathbb{1}_{z=z_i}$

$$\langle p(z) \rangle = \mathop{\mathbb{E}}_{z_i} p(z) = \frac{1}{Z} \sum_{\{z_i\}} p(z)\, e^{-\beta E(z_i)}$$

$$= \frac{1}{Z} \sum_{\{z_i\}} \frac{1}{N} \mathbb{1}(z - z_i)\, e^{-\beta \sum z_i}$$

$$= \frac{N}{Z} \sum_{z_i} \frac{1}{N} \mathbb{1}(z - z_i) e^{-\beta z_i} \underbrace{\sum_{z_j \neq z_i} e^{-\beta \sum z_j}}_{\left(\frac{1}{1-e^{-\beta}}\right)^{N-1}}$$

$$= (1-e^{-\beta})\, e^{-\beta z}$$

$$p(z) = \langle p(z) \rangle =: p^{eq}(z)$$

For a snapshot of N particles the probability of seeing a given $p(z) = p_\lambda(z)$ goes as

$$\exp[-N D_{KL}(p_\lambda \| p^{eq})]$$

For a more general potential: $p_{eq} = \langle p(z) \rangle = \frac{e^{-\beta V(z)}}{Z(\beta)}$

$$\Rightarrow D_{KL}(p \| p_{eq}) = \beta \sum_x p(x) V(x) + \sum_x p(x) \log p(x) + \log Z(\beta)$$

Lets say we want

$\mathbb{E}[F]$ through $\hat{f} = \frac{1}{N} \sum_{i=1}^{N} F(s_i)$

For $A \subset \mathbb{R}$ an interval

$$\mathbb{P}_{\underline{s}}[\hat{f} \in A] \doteq \exp[-N\, I(A)]$$

$$I(A) = \min_q \left[D(q \| p) \;\middle|\; \sum q(x) F(x) \in A \right]$$

Theorem: (Sanov) $\underline{s} = s_1 \cdots s_N$ iid $\sim p(x)$
if q is empirical for $\underline{s}$ then for any compact subset K of distributions on $\mathcal{X}$

$$\Pr_{\underline{s}}[q \in K] \doteq \exp[-N\, D_{KL}(q^* \| p)]$$

$$\boxed{q_* = \operatorname*{argmin}_{q \in K} D_{KL}(q \| p)}$$

Now, back to $\hat{f}$. We have $\hat{f} = \mathbb{E}_{q_{\underline{s}}}[F]$

Then use $K_{\hat{f} \in A} = \{ q \mid \mathbb{E}_q[F] \in A \}$

$$\Rightarrow \quad \mathbb{P}_{\underline{s}}[\hat{f} \in A] \doteq \exp[-N \min_{q \in K} D_{KL}(q \| p)]$$

Consequence: $s_1 \cdots s_N \sim p$ w/ bounded support

$$\mathbb{P}[s_1 + \cdots + s_N \leq 0] \doteq \exp[-N \inf_q D_{KL}(q \| p)]$$

Eg 1:

$A = (-\infty, 0)$

$K = \{ q \mid \mathbb{E}_q[x] \le 0 \}$ $\qquad \exp\left[-\inf_{\substack{q \\ \text{s.t. } \mathbb{E}_q[x] \le 0}} D_{KL}(q \| p) \right]$

KKT

$$\sum_x \left(q(x) \log \frac{q(x)}{p(x)} - \lambda x q(x) - \lambda' q(x) \right)$$

$$\frac{\delta}{\delta q(x)} \to 1 + \log \frac{q(x)}{p(x)} - \lambda x - \lambda' = 0$$

$$\Rightarrow \boxed{q = \frac{p(x) e^{-\lambda x}}{Z(\lambda)}} \Rightarrow D_{KL}(q \| p) = \int \frac{p(x) e^{-\lambda x}}{Z(\lambda)} \left[-\lambda x - \log Z \right]$$

$$= \mathbb{E}_q[-\lambda x] - \log Z$$

$$\Rightarrow \exp[-D_{KL}] = \mathbb{E}_{x \sim p}\left[e^{-\lambda x} \right] + \lambda \, \mathbb{E}_{x \sim p}\left[e^{-\lambda x} x \right] \Big/ \left(\mathbb{E}_{x \sim p}\left[e^{-\lambda x} \right] \right)^{-1}$$

$\uparrow \ge 0$ $\qquad \uparrow$ why subleading?

$$\Rightarrow \Pr[S_1 + \cdots S_N \le 0] \doteq \left\{ \inf_{\lambda \ge 0} \mathbb{E}_{s \sim p} e^{-\lambda s} \right\}^N$$

N.B. when $\mathbb{E}_p[x] \le 0$ this is just 1 as $N \to \infty$

Eg 2 For N particles w/ $E(\underline{z}) = \sum z_i$

we get $\bar{z} = \frac{1}{N} \sum z_i$

has $z_{eq} = \mathbb{E}(\bar{z}) = \frac{1}{N} \frac{\partial}{\partial \beta} \log Z(\beta) = \frac{e^{-\beta}}{1 - e^{-\beta}} = \frac{1}{e^{\beta} - 1}$

$$P_{\underline{z}}(\bar{z} > z) \doteq \exp[-N I(z)]$$

$$I(z) = \inf_{q \text{ s.t. } \underset{z\sim q}{E}[z] > z} D_{KL}(q \| p)$$

$$q(x) = \frac{p(x)\, e^{-\lambda x}}{Z(\lambda)} = (1-e^{-\lambda})\, e^{-\lambda x}$$ s.t.

different λ

$$\underset{\lambda}{E}[z] = (e^{\lambda}-1)^{-1} =: z \Rightarrow \lambda = \log(1+z^{-1})$$

$$I(z) = D_{KL}(p_\lambda \| p_{eq}) = \log \frac{1-e^{-\lambda}}{1-e^{-\beta}} + \frac{\beta-\lambda}{e^{\lambda}-1}$$

$$= \log \frac{1+z_{eq}}{1+z} + z\left(\log \frac{1+z_{eq}}{1+z} + \log \frac{z}{z_{eq}}\right)$$

$$= (1+z) \log \frac{1+z_{eq}}{1+z} + z \log \frac{z}{z_{eq}}$$

$$\exp[-N I(z)] \doteq \Pr_{\bar{z}}[\bar{z} > z]$$

$$\doteq \Pr_{\bar{z}}[\bar{z} < z]$$

Exercise

Build a thermometer

1) Take a snapshot of all N particles

$\Rightarrow$ get $\bar{z}$, take $\hat{\beta} = \log(1+z^{-1})$

$$\Pr_{z_1 \cdots z_n \sim p}[\hat{\beta} > \beta] = \exp\left[-N \inf_{\substack{q_* \text{ s.t.} \\ \underset{q}{E} \log(1+z^{-1}) > \beta}} D_{KL}(q \| p)\right]$$

$$\mathcal{L} = q \log \frac{q}{p} - \lambda q \log(1+z^{-1}) - \lambda' q$$

$$\Rightarrow 1 + \log \frac{q}{p} - \lambda \log(1+z^{-1}) - \lambda'$$

$$\Rightarrow q \sim p(z)\,(1+z^{-1})^{\lambda}$$ Take $\bar{x} \Rightarrow \hat{\beta} = \log(1+\bar{x}^{-1})$

$$E[\bar{x}] = (e^{\beta}-1)^{-1}$$

4.2.3 Asymptotic Equipartition

Prop 4.7 # of sequences of type belonging to $K \in \mathcal{M}(\mathcal{X})$ goes as

just counting

$$N_{K,N} \doteq \exp[N\, H(q^*)]$$

$$q^* = \arg\max_q H(q)$$

Proof

Take reference distribution $p(x)$ to be uniform

$$N_{K,N} = |\mathcal{X}|^N \, \mathbb{P}_{\underline{s} \sim p}[q_{\underline{s}} \in K]$$

Sanov $$\doteq \exp[N \log |\mathcal{X}| - N\, D_{KL}(q_* \| \text{unif})]$$

$$= \exp[N\, H_{q_*}]$$

For a sequence $\underline{s}$ $\quad r(\underline{s}) := -\frac{1}{N} \log P_N(s)$ (joint)

$$= -\frac{1}{N} \sum_i \log p(s_i)$$

$$\mathbb{E}_{s_i}\, r(s) = H(p) \qquad = \mathbb{E}_{q_{\underline{s}}} \log \frac{q_{\underline{s}}}{p_{\underline{s}}} + H(q)$$

$\underline{s}$ is ε-typical if $|r(s) - H(p)| < \varepsilon$

$T_{N,\varepsilon} :=$ ε-typical sequences

Theorem

✓ 1) $\lim_{N\to\infty} \Pr[\underline{s} \in T_{N,\epsilon}] = 1$

2) For N large

$$\exp[N(H(p)-\epsilon)] < T_{N,\epsilon} < \exp[N(H(p)+\epsilon)]$$

Def'n of $T_{N,\epsilon}$ 3) For any $\underline{s} \in T_{N,\epsilon}$ $e^{-N(H(p)+\epsilon)} < P_N(\underline{s}) < e^{-N(H(p)-\epsilon)}$

For 1) use atypicality:

$$\Pr[\underline{s} \notin T_{N,\epsilon}] \sim \exp[-N \min_{\text{s.t. } |H(q)-H(p)| \geq \epsilon} D(q\|p)]$$

finite ball $\Rightarrow D_{KL} > 0$ with a gap not going to zero w/ N

For 2) ϵ-typical is the same as

$$|D_{KL}(q\|p) + H(q) - H(p)| < \epsilon$$

>0 not if $H(p) > H(q)$

$\Rightarrow |H(q) - H(p)| < \epsilon$ why?

$$\Rightarrow |T_{N,\epsilon}| \doteq \exp[N H(q)] \doteq \exp[N(H(p) \pm \epsilon)]$$

For 3) Def'n of $T_{N,\epsilon}$

4.3 Correlated Variables

$$P(\underline{x}) = P_N(x_1 \cdots x_N)$$

want $\mathbb{E}\, \mathcal{F}(\underline{x})$

$$\mathcal{F}(\underline{x}) := \frac{1}{N} \sum_i f(x_i)$$

First assume $\mathcal{F}$ follows a large-deviation principle

$$P_N(\mathcal{F}) \doteq \exp[-N I(\mathcal{F})]$$

Define $\psi_N(t) = \frac{1}{N} \log \mathbb{E}\, e^{Nt\bar{F}}$

chemical potential for F

$$\lim_{N\to\infty} \psi_N(t) \sim \frac{1}{N} \log \int d\bar{F} \exp[Nt\bar{F} - NI\bar{F}]$$
$$=: \psi(t)$$

$$\psi(t) = \sup_{\bar{F}} \; t\bar{F} - I(\bar{F})$$

convex in t

$$I_\psi(\bar{F}) = \sup_{t} \; t\bar{F} - \psi(t)$$

convex envelope of $\bar{F}$

Eg 1) Ising N spins $P_N[\sigma] = \frac{e^{-\beta E(\sigma)}}{Z(\beta)}$

want: $m(\sigma) = \frac{1}{N}\sum_i \sigma_i$

$$\log \mathbb{E}\, e^{Nt\,m(\sigma)} = \log Z(\beta, B = \tfrac{t}{\beta}) - \log Z(\beta, 0)$$

We can get ψ explicitly $= \log \lambda^{T}_{max}$

$$\Rightarrow \psi(t) = \log\left(\frac{\cosh t + \sqrt{\sinh^2 t + e^{-4\beta}}}{1 + e^{-2\beta}}\right)$$

Can then numerically get $I_\psi(t)$

Eg 2) Markov chain $w(x \to y) > 0$

Define $w_t(x \to y) = w(x \to y) \exp(t F(y))$

eigenvalue problem $\sum_x \varphi(x)\, w(x \to y) = \lambda(t)\, \varphi(y)$

Take $\lambda(t) = \lambda^{max}(t)$ (unique by Perron-Frobenius) (Positive matrix has unique λ^{max} and corresp. positive vec)

$\Rightarrow \Psi(t) = \log \lambda(t)$

1D ising is a special case?

T-matrix turned it into a 2-state markov chain.

Eg 3) Curie-Weiss $E(\underline{\sigma}) = -\frac{1}{N} \sum_{(ij)} \sigma_i \sigma_j$

look at $\bar{m}(\sigma) = \frac{1}{N} \sum_i \sigma_i \in [-1, 1]$

$E(m) = \frac{N}{2}(m^2 - 1) \qquad S(m) = \log \binom{N}{N\frac{(1+m)}{2}} \doteq N\, \mathcal{H}\left(\frac{1+m}{2}\right)$

$$\mathbb{P}_N\left[\bar{m} \in [m_1, m_2]\right] \doteq \frac{1}{Z_N(\beta)} \int_{m_1}^{m_2} e^{N \varphi_{mF}(m, \beta)}\, dm$$

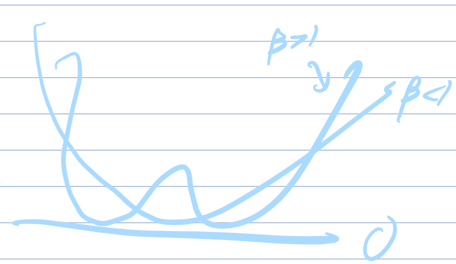
$\varphi_{mF}(m; \beta) = \underbrace{\frac{\beta m^2}{2}}_{E} - \log \underbrace{2 \cosh \beta m}_{e^{\beta m} + e^{-\beta m}}$ ← same EOM ⇒ good enough

$m = \tanh \beta m$

$I(m) = \varphi(m^*, \beta) - \varphi(m, \beta)$

saddle point value from $Z(\beta)$ in denom

$\Rightarrow I(m) \sim$



Now $\Psi(t) = \frac{1}{N} \log \mathbb{E}\, e^{tm}$

$\to \sup_m \; tm - I(m)$

magnetic field

For $\beta < 1$ $\quad \Psi(t)$ is convex

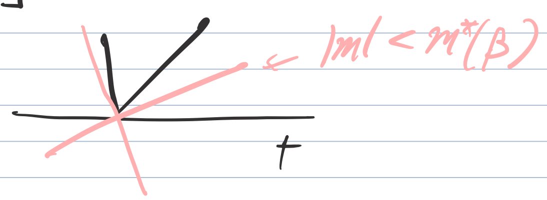
For $\beta > 1$ $\quad \Psi(0)$ is discontinuous

$\Psi'(0+) = m^*(\beta) = -\Psi'(0^-)$

$\Rightarrow$ For $m \in [-m_*(\beta), m_*(\beta)]$

$\sup_t \; mt - \Psi(t)$

is at $t=0$



$\Rightarrow I_\Psi(m) = 0$

otherwise its at unique sol'n $\Psi'(t) = m$

$\Rightarrow I_\Psi$ is convex envelope of m

$\Rightarrow$ estimating $P_N(m) \doteq \exp[-N I_\Psi(m)]$

would overestimate the probabilities of $m \in (-m^*(\beta), m^*(\beta))$

Gartner-Ellis Theorem

Take $\xi(\underline{x})$ and assume the moment generating fn

$$\Psi_N(t) := \frac{1}{N} \log \mathbb{E}[e^{t\xi}]$$

exists as $N \to \infty$

Def $I_\Psi(\xi) = \sup_t \; \xi t - \Psi(t)$

Let $\mathcal{E}$ be the set of points where I_Ψ is C^2, $I'' > 0$

1) For any closed set $F \subseteq \mathbb{R}$

$$\limsup_{N \to \infty} \frac{1}{N} \log P_N(\xi \in F) \le -\inf_{\xi \in F} I_\Psi(\xi)$$

2) For any open set $G \subseteq \mathbb{R}$

$$\lim_{N\to\infty} \sup \frac{1}{N} \log P_N(\mathfrak{f} \in G) \leq -\inf_{\mathfrak{f} \in G \cap \mathcal{E}} I_\psi(\mathfrak{f})$$

3) If $\psi(t)$ is always differentiable can take $G \cap \mathcal{E} \to G$

4.3.3 Typical Sequence

Take again $r(\underline{x}) = -\frac{1}{N} \log P(\underline{x})$

$$\Rightarrow \psi_N(t) = -\frac{1}{N} \log \sum_{\underline{x}} P_N(\underline{x})^{1-t}$$

Take $P_N(\underline{x}) = \dfrac{\exp(-\beta E_N(\underline{x}))}{Z_N(\beta)}$

$$\Rightarrow \psi_N(t) = (1-t)\beta F_N(\beta) - \beta F_N((1-t)\beta)$$

$$\beta F_N(\beta) = -\frac{1}{N} \log Z(\beta)$$

assume $F_N(\beta)$ is finite as $N \to \infty$

$$\Rightarrow P_N(\mathfrak{f} \in F) \doteq \exp\left[-N \inf_{\mathfrak{f} \in F} I_\psi(F)\right]$$

if $\mathfrak{f}$ is discontinuous, $r(\underline{x})$ may take several values with non-vanishing probability

N.B. I can't use $I(\mathfrak{f})$ over $I_\psi(\mathfrak{f})$ because of domain wall formation making $m \in [-m^*, m]$ more likely than just $\exp[-N\, I(m)]$

4.4 Gibbs Free Energy

Boltzmann has $\mu_\beta = \exp[-\beta(E(x) - F(\beta))]$

$$G_\beta[P] := \sum_{x \in \mathcal{X}} P(x) E(x) + \frac{1}{\beta} \sum P(x) \log P(x)$$

$$= \frac{1}{\beta} D_{KL}(P \| \mu_\beta) + F(\beta)$$

Unique minimum of G_β is at $P = \mu_\beta$

Then $G_\beta = F(\beta)$

Probability of empirical dist $P(\underline{x})$ being far from $\mu_\beta(\underline{x})$ is

$$\mathbb{P}[P] \doteq \exp[-N(G_\beta(P) - F(\beta))]$$

If I can't compute Z (therefore F)

lets instead minimize G over a trial subspace

Eg 1 $\quad E(x) = \frac{1}{2} t x^2 + \frac{1}{4} x^4$

$\beta = 1 \quad F(t) = -\log \int dx\, e^{-E(x)}$

Trial dist: $\quad Q_a = \frac{1}{\sqrt{2\pi a}} e^{-x^2/2a}$

$$\Rightarrow G[Q_a] = \frac{1}{2} t a + \frac{3}{4} a^2 - \frac{1}{2}(1 + \log 2\pi a)$$

Eg 2 $\quad$ Same problem with $Q_a = \frac{1}{\sqrt{2\pi}} e^{-(x-a)^2/2}$

$$G[Q_a] = a^4 + 2a^2(3+t) + \frac{3}{4} + \frac{t}{2} - \frac{1}{2}(1 + \log 2\pi)$$

$$\Rightarrow a^3 = -2a(3+t)$$

so $a = 0 \quad (t > -3) \quad \Rightarrow G = \text{const}$

or $a = \pm\sqrt{-(3+t)}$

two ground states

4.4.2 Mean Field

Take Ising $\quad E(\underline{\sigma}) = -\frac{1}{2d} \sum_{(ij)} \sigma_i \sigma_j - B \sum_i \sigma_i$

d links per site

take $Q_m = \prod_i q_m(\sigma_i)$

$$q_m(\pm 1) = \frac{1 \pm m}{2} \quad \Rightarrow \mathbb{E}_{Q_m}[\sigma_i] = m$$

$$G[Q_m] = \sum_{\underline{\sigma}} Q_m(\underline{\sigma}) E(\underline{\sigma}) - \frac{1}{\beta} \mathcal{H}(Q_m)$$

$$= -\frac{1}{2} m^2 - BNm - \frac{N}{\beta} \mathcal{H}\left(\frac{1+m}{2}\right)$$

$$g(m;\beta,B) := G[Q_\beta]/L^d = -\frac{1}{2}m^2 - Bm - \frac{1}{\beta}\mathcal{H}\left(\frac{1+m}{2}\right)$$

4.5 Monte Carlo

Sampling $\underline{x} \in \mathcal{X}^N$ from $P(\underline{x})$ is challenging at large N

Use MCMC: $w(x \to y)$

1) irreducible: $\forall x,y \quad w(\underline{x} \to \underline{y}) > 0$
2) aperiodic: $w(\underline{x} \to \underline{x}) > 0$
3) $P(\underline{x})$ is stationary for w

$$\sum_x P(x)\, w(x \to y) = P(y)$$

stronger version is detailed balance / reversibility:

$$P(x)\, w(x \to y) = P(y)\, w(y \to x)$$

Theorem: Assuming 1-3, let $\underline{X}_t$ be random vars distributed according to MC $w(\underline{x} \to \underline{y})$

$\underline{X}_0 = \underline{x}_0$

Then:

1) $\lim_{t \to \infty} \mathbb{P}[\underline{X}_t = x] = P(\underline{x})$

2) $\lim_{t \to \infty} \frac{1}{t}\sum_{s=1}^{t} f(\underline{X}_s) = \mathop{\mathbb{E}}_{P} f(\underline{x})$

Motivating Example

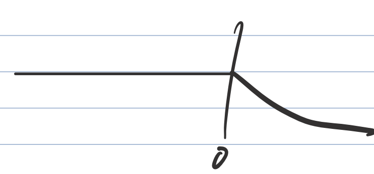
N Ising spins $\underline{\sigma}$

Sample $\mu_\beta(\underline{\sigma})$. Def $\underline{\sigma}^{(i)} = \underline{\sigma}$ except $\sigma_i \to -\sigma_i$

$$\Delta E^{(i)} = E(\underline{\sigma}^{(i)}) - E(\underline{\sigma})$$

At each step choose i randomly

$$w_i(\underline{\sigma}) = \exp[-\beta\, [\Delta E_i]_+]$$



$$w(\underline{\sigma} \to \underline{\tau}) = \frac{1}{N} \sum_i w_i(\underline{\sigma}) \, \mathbb{1}(\underline{\tau} = \underline{\sigma}^{(i)}) + \left[1 - \sum_i w_i(\underline{\sigma})\right] \mathbb{1}(\underline{\tau} = \underline{\sigma})$$

$$\mu_\beta(\underline{\sigma}) \, w_i(\underline{\sigma}) = \frac{\exp}{Z}\left[-\beta\left(E(\underline{\sigma}) + [\Delta E_i]_+\right)\right] = \frac{\exp}{Z}\left[-\beta\left(E(\underline{\sigma}^{(i)}) + [-\Delta E_i]_+\right)\right]$$

$E(\underline{\sigma})$ if $\Delta E_i < 0$
$E(\underline{\sigma}^i)$ if $\Delta E_i > 0$

$$= w_i(\underline{\sigma}^{(i)}) \, \mu_\beta(\sigma^{(i)})$$

Exercise: Heat bath algorithm (ie glauber)

$$w_i(\underline{\sigma}) = \frac{1}{2}\left[1 - \tanh\left(\frac{\beta \Delta E}{2}\right)\right] = \frac{1}{1 + e^{\beta \Delta E}}$$

$$\mu_\beta(\underline{\sigma}^{(i)}) \, w_i(\underline{\sigma}^{(i)} \to \underline{\sigma}) = \frac{1}{Z} \frac{e^{-\beta E(\sigma^{(i)})}}{1 + e^{\beta(E(\underline{\sigma}) - E(\underline{\sigma}^i))}}$$

$$= \frac{1}{Z} \frac{1}{e^{\beta E(\underline{\sigma})} + e^{\beta E(\underline{\sigma}^i)}}$$

$$= \frac{e^{-\beta E(\underline{\sigma})}}{Z} \frac{1}{1 + e^{\beta(E(\sigma^i) - E(\sigma))}}$$

$$= \mu_\beta(\underline{\sigma}) \, w(\underline{\sigma} \to \underline{\sigma}^i)$$

Can generalize to K-state systems

- Draw i from $1 \to N$

$$x_{-i}^t = x_{-i}^{t-1}$$
$$x_i^t \sim \mathbb{P}\left[X_i = k \mid \underline{X}_{-i} = x_{-i}^{t-1}\right] = \text{softmax}(\beta \Delta E_k)$$

1) 2) 3) Follow

Timescales in MCMC

MCMC @ time t

1) let $d_{x_0}(t) = \| P_t(\cdot \mid x_0) - P(\cdot) \|_1$

$\tau_{eq}(\varepsilon) = \min \{ t : \sup_{x_0} d_{x_0}(t) < \varepsilon \}$ exponential convergence

usually $\varepsilon = 1/e$

2) Say MCMC is burned in

$$\bar{O}_T := \frac{1}{T} \sum_{t=0}^{T-1} O(\underline{X}_t)$$

$$\langle \bar{O}_T \rangle = \mathbb{E}_P[O]$$

$O_t := \bar{O}(X_t)$

$$\mathrm{Var}\, \bar{O}_T = \frac{1}{T^2} \sum_{s,t} \langle O_s \; O_t \rangle_c$$

$$= \frac{1}{T^2} \sum_t (T-t) \langle O_0 \; O_t \rangle_c$$

$$C_O(t-s) = \frac{\langle O_t \; O_s \rangle_c}{\langle O_0 \; O_0 \rangle_c}$$

$$\Rightarrow \mathrm{Var}\, \bar{O}_T = \frac{\langle O_0 \; O_0 \rangle_c}{T^2} \sum_{t=0}^{T} (T-t)\, C_O(t)$$

On finite state space

$$\mathrm{Var}\, O_T = \frac{\tau^O}{T} \left(\mathrm{Var}_P\, O \right) + O(T^{-2})$$

$$\tau^O := \sum C_O(t)$$

4.6 Simulated Annealing

Optimization recovered as $\beta \to \infty$

Taking random init + $\beta \to \infty$ "quench"

corresponds to greedy algorithm

since $w(i \to j)$ looks like 

$\Rightarrow \beta \to \infty$ MCMC is not irreducible

Alt: β large but finite

How large should it be? How long should we wait?

Assume WLOG min @ $E_0 = 0$

$$\Pr_{\mu_\beta}[E=0] \Rightarrow$$

$$\psi_N = \frac{1}{N} \log \sum_x \mu_\beta e^{tE(x)} = \frac{1}{N} \log \frac{\sum_x e^{-(\beta - t)E(x)}}{\sum_x e^{-\beta E(x)}}$$

$$\boxed{\psi_N = \varphi_N(\beta - t) - \varphi_N(t)}$$

$$\mu_\beta[E=0] \doteq \exp[N \psi_N(-\infty)] = 1/\tau$$

$$= \exp[N(\varphi(\infty) - \varphi(\beta))]$$

$\Rightarrow$ must wait $\tau = \exp[N(\varphi_N(\beta) - \varphi_N(\infty))]$

1) if φ_N has well defined limit as $N \to \infty$

we must scale $\beta \to \infty$ with $N \to \infty$

for τ to be finite

2) However as $\beta \to \infty$ MC equilibration time diverges

$\Rightarrow$ let β vary over t in MCMC

"annealing schedule"

ie time-dep MCMC

$\Rightarrow$ "Avoids both of the above problems"

Can indeed work in practice

4.7 Sonov's Theorem via physics

$$\mathbb{P}_{x_i \sim p}[q(x)] = \mathbb{E}_{x_i \sim p} \prod_x \mathbb{1}\left[q(x) = \frac{1}{N}\sum_{i=1}^N \delta_{x, x_i}\right]$$

functional of q

agreement on all x

$$= \sum_{x_1 \cdots x_N} p(x_1)\cdots p(x_N) \underbrace{\prod_x \mathbb{1}\left[q(x) = \frac{1}{N}\sum_{i=1}^N \delta_{x=x_i}\right]}_{\int_0^{2\pi} \frac{D\lambda(x)}{(2\pi)^N} \exp\left[i\langle \lambda, Nq(x) - \sum_{i=1}^N \delta_{x,x_i}\rangle\right]}$$

$$\Rightarrow \mathbb{P}[q(x)] \propto \int D\lambda(x) \exp\left[iN\langle \lambda, q\rangle\right] \prod_{i=1}^N \left(\sum_{x_i} p(x_i) \exp[-i\lambda(x_i)]\right)$$

$$= \int D\lambda \exp[iN\, \lambda \cdot q] \langle p, e^{-i\lambda}\rangle^N$$

$$\Rightarrow \mathbb{P}[q(x)] \propto \int D\lambda(x) \exp[N S(\lambda)]$$

$$S(\lambda) = i\langle \lambda, q\rangle + \log \langle p, e^{-i\lambda}\rangle$$

$\Rightarrow$ Saddle $\quad \frac{\delta S}{\delta \lambda} = 0 \Rightarrow q(x) = \frac{p(x)\, e^{-i\lambda(x)}}{\underbrace{\sum_{x'} p(x') e^{-i\lambda(x')}}_{\text{const}}}$

$$\lambda = i \log \frac{q}{p} + C$$

Free so $\sum p = 1$

$$\Rightarrow S[\lambda^*] = -D_{KL}[q \| p] + S_0$$