log P(x) are log-posterior likelihoods or log-Bayes factors. They will

\[
\log P(x) = \log \frac{P(x)}{q(x)} = \log \frac{P(x)}{Q(x)} = \log \frac{P(x) Q(x)}{Q(x) Q(x)} = \log \frac{P(x) Q(x)}{Q(x) Q(x)} - \log \frac{Q(x)}{Q(x)}
\]

\[
\text{log-likelihood} = \log P(x) - \log q(x)
\]

\[
\text{MRF INFERENCE: (FINISHING) VARIATIONAL INFERENCE}
\]

\[
\Pi_q(x) = \prod q_i(x_i)
\]

\[
P(x) = \frac{1}{Z} \exp \sum_i H_i(x_i)
\]

\[
Q(x) = \prod_i q_i(x_i)
\]
play a very important role in MAP Inference later, where it is common to always work in log-domain.

Notation \( \Theta_c(x_c) = \log \Phi_c(x_c) \)

So in a pairwise MRF \( \Theta_i(x_i) = \log \Phi_i(x_i) \) 
\& \( \Theta_{ij}(x_i, x_j) = \log \Phi_{ij}(x_i, x_j) \)

Node to edge "Scores"

In some papers, people also use negative log-potentials or "Energies". All algorithms have equivalent formulations in "Scores" & "Energies".

\[
S(X) = \sum_c \Theta_c(x_c) = \sum_c \log \Phi_c(x_c)
\]
\[
E(X) = \sum_c E_c(x_c) = \sum_c -\log \Phi_c(x_c)
\]

Back to team #2

\[
= \prod \sum_{x_c} q_c(x_c) \Theta_c(x_c) - \log Z
\]

\[
= \prod \sum_c \sum_{x_c} q_c(x_c) \Theta_c(x_c) - \log Z
\]

\[
= \prod \sum_c q_c(x_c) \Theta_c(x_c) - \log Z
\]

where \( q_c(x_c) = \prod q_i(x_i) \)
Putting it all together

\[ \text{KL}(q \parallel p) = -H_q(X) - \sum_c \sum_{x_c} q_c(x_c) \Theta_c(x_c) + \log Z \]

\[ \Rightarrow \log Z = \text{KL}(q \parallel p) + H_q(X) + \sum_c \sum_{x_c} q_c(x_c) \Theta_c(x_c) \]

Called "Gibbs Free Energy"

\[ F[p, q] \]

\[ \Rightarrow \log Z = \text{KL}(q \parallel p) + F[p, q] \]

Note 1: \( \log Z \) is constant w.r.t \( q \)

\[ \Rightarrow \arg \min_q \text{KL}(q \parallel p) = \arg \max_q F[p, q] \]

Note 2: \( \text{KL}(q \parallel p) \geq 0 \) [why? see HMC in Intro to ML]

\[ \Rightarrow \log Z \geq F[p, q] \]

Ah, so we're maximizing a lower bound on \( \log Z \)!

Note 3: \( \theta \) Maximizing \( F[p, q] \) "makes sense"

\[ F[p, q] = H_q(X) + \sum_c \sum_{x_c} q_c(x_c) \Theta_c(x_c) \]

\[ \frac{\text{Eq}[\Theta_c(X)]}{\text{Eq}[\Theta_c(X)]} \]

So find me a simple \( q \) that puts heavy mass on high scoring configurations AND is as uncertain as possible!
Note 4: There will usually be a trade-off

If $q$ is not simple enough, we can always get $\text{KL}(q||p) = 0$ by setting $q = p$
so lower bound is tight
but can't perform $\max_q F[p,q]$ with complicated distributions!

2) Mean-Field Update Rules

So how do we maximize $F[p,q]$?
Let's make the problem:

$$\max_{q_i} \sum_i q_i(x_i) \log q_i(x_i) + \sum_i \sum_{x_i} q_i(x_i) \theta_i(x_i)$$

s.t. $\sum_{x_i} q_i(x_i) = 1$ : $x_i \forall i$

$q_i(x_i) \geq 0$

Lagrangian Multiplier Method!

[Assume pairwise MRF for simplicity, derivation generalizes to arbitrary MRFs]

$\Rightarrow \sum_{x_i} q_i(x_i) \theta_i(x_i) = \sum_{x_i} q_i(x_i) \Theta_i(x_i)$

$+ \sum_{i,j} q_i(x_i) q_j(x_j) \Theta_{ij}(x_i, x_j)$
\[ L(q_i(x), \{x_i\}) = -\frac{1}{2} \sum_{i} q_i(x_i) \log q_i(x_i) \]
\[ + \sum_{i} q_i(x_i) \Theta(x_i) \]
\[ + \sum_{j} \sum_{i} q_i(x_i) q_j(x_j) \Theta_{ij}(x_i, x_j) \]
\[ - \sum_{i} x_i \left( \frac{1}{2} q_i(x_i) - 1 \right) \]

Now, \( \frac{\partial L}{\partial q_s(x_s)} = 0 \) \([\text{to find stationary point}]\)

\[ \Rightarrow \quad - \frac{q_s(x_s)}{q_s(x_s)} - \log q_s(x_s) \]
\[ + \Theta_s(x_s) \]
\[ + \sum_{t \in \text{enlarge}} \sum_{x_t} q_t(x_t) \Theta_{st}(x_s, x_t) \]
\[ - x_s \]

\[ \Rightarrow \quad \log q_s(x_s) = -1 - x_s + \Theta_s(x_s) + \sum_{t \in \text{enlarge}} \sum_{x_t} q_t(x_t) \Theta_{st}(x_s, x_t) \]

\[ \Rightarrow \quad q_s(x_s) = e^{-1-x_s} e^\Theta_s(x_s) \prod_{t \in \text{enlarge}} e^{\sum_{x_t} \Theta_{st}(x_s, x_t) q_t(x_t)} \]
\[ q_s(x_s) = \frac{1}{Z_s} \psi_s(x_s) \prod_{t \in \text{tree}} e^{\sum_{x_t} \Theta_s(x_s, x_t) q_t(x_t)} \]

This is a fixed point equation. Any \( q \) satisfying this is a fixed point.

How do we reach such a \( q \)?

Iterative update rules:

\[ q_{s}^{(t+1)}(x_s) = \frac{1}{Z_s} \psi_s(x_s) \prod_{t \in \text{tree}} e^{\sum_{x_t} \Theta_s(x_s, x_t) q_t^{(t)}(x_t)} \]

This is guaranteed to converge to a fixed point. But that fixed point may not be the optimum maxima! Non-convex objective.

Update rule looks like BP but isn't.

Need a different assumption on \( q \) to get BP updates!

[Advanced topic: won't cover in class?]

Maybe Final? "E"