Motivation

Modeling gene alterations

Mutual exclusivity

Co-relevance

Modeling teams

Ground set: $V = \{1, \ldots, n\}$

Data: $D = \{S\}_S \subseteq V$

Model higher-order interactions

$\eta(S; \theta) = \frac{1}{Z(\theta)} \exp(F(S; \theta))$

- Directed interactions
- Log-Metropolis
- Log-Model

When Gibbs fails

$\Omega_1$

$\Omega_2$

Sampling and Inference

Learning

Gibbs sampling

Max. Likelihood

Approximate $\nabla_\theta \mathcal{L}(\theta)$

Exact computation : $\text{NP-hard in general}$

Sample from $p(\theta | D)$

Contributions

- We propose the $M^1$ sampler, which makes global moves to avoid bottlenecks.
- For a specific class of log-models, we show that combining Gibbs and $M^1$ results in an exponential mixing time improvement over Gibbs.
- We propose a semigradient-based mixture construction and demonstrate its effectiveness on three models learned from real-world data.

The $M^1$ Chain

$M^1 = \text{Mixture of Log-Metropolis}$

\[ \eta(S; T) = \eta(T) = \frac{1}{Z(\theta)} \sum_{m(S)} \exp(m(S; \theta)) \]

$M^1 = \{\eta(S; \theta) = \eta(T)\}$

- Target: $\eta(S) \propto \exp(F(S))$
- Proposal: $\eta(T)$
- Accept with probability $\min\left\{ 1, \frac{\eta(T)}{\eta(S)} \right\}$

Proposition 1

Mixture $\eta$ can approximate any distribution $p$ arbitrarily well.

BUT may need an exponential $(n \times m)$ number of components $r$

The Combined Chain

Gibbs step with prob. $\alpha$

$M^1$ step with prob. $1 - \alpha$

Decomposition theorem [Jerrum et al., '04]

Projection chain

Gibbs

Restriction chains

Class of log models on the complete graph (Cone-Weis)

Experiments

PSRF


SemiGradient

Ideally would want to minimize

For $i = 1 \text{ to } d$

$\sigma_i = \text{Perm}(V)$

$m_{\sigma_i} = \text{SemiGradient}(F; \sigma_i)$

return $(m_{\sigma_1}, \ldots, m_{\sigma_d})$

$\text{SemiGradient}$

- Submodularity $\Rightarrow$ natural bisecting property

- Sublinear $\Rightarrow$ special bisecting property

- Construction works for general set function $F$