Semidefinite relaxations for approximate inference on graphs with cycles

Speaker: Martin Wainwright, EECS, UC Berkeley

`wainwrig@eecs.berkeley.edu`

Joint work with: Michael Jordan, CS & Statistics, UC Berkeley

`jordan@cs.berkeley.edu`

Introduction

- graphical models are used and studied in various fields
  (e.g., machine learning; error-correcting coding; statistical physics; computer vision)

- following problems are important but difficult:
  (a) computing marginal distributions
  (b) estimating model parameters from data

- role of variational methods
  (a) mean field methods (e.g., Jordan et al., 1999)
  (b) Bethe/Kikuchi approximations and variations
  (e.g., Yedidia et al., 2001; Minka, 2001; McEliece & Yildirim, 2002, Pakzad & Anantharam, 2002)

Overview

Possible concerns with the Bethe/Kikuchi problems and variations?

(a) lack of convexity $\Rightarrow$ multiple local optima, and substantial algorithmic complications

(b) failure to bound the log partition function

Overview

Possible concerns with the Bethe/Kikuchi problems and variations?

(a) lack of convexity $\Rightarrow$ multiple local optima, and substantial algorithmic complications

(b) failure to bound the log partition function

Goal: Techniques for approximate inference and parameter estimation based on:

(a) convex variational problems $\Rightarrow$ unique global optimum

(b) relaxations of exact problem $\Rightarrow$ upper bound on log partition function

Variational approach

Basic idea: Represent a quantity of interest $\widehat{z}$ as the solution of an optimization problem:

(a) study $\widehat{z}$ via the optimization problem.

(b) approximate $\widehat{z}$ by approximating the optimization problem.

Goal: Obtain a variational representation for:

(a) the log partition function.

(b) the inference problem of computing $\mu_\alpha := \sum_{\mathbf{x}} p(\mathbf{x})\phi_\alpha(\mathbf{x})$.

Classical form of convex duality

- let $\mathcal{P}$ be the set of all possible distributions over $\mathbf{x}$

- log partition function can be recovered as a maximum entropy problem over $\mathcal{P}$:

$$\log Z_p \quad = \quad \max_{q \in \mathcal{P}} \Big\{ \sum_{\mathbf{x}} q(\mathbf{x}) \big[\sum_{\alpha} \theta_\alpha \phi_\alpha(\mathbf{x}) \big] \quad + \quad H(q) \Big\}$$

  where H is the usual (Boltzmann-Shannon) entropy

$$H(q) \quad := \quad - \sum_{\mathbf{x} \in \mathcal{X}^n} q(\mathbf{x}) \log q(\mathbf{x}).$$

- equivalent to the assertion $\min_{q \in \mathcal{P}} D(q \parallel p) = 0$.

Exponential representations

Parameterized family of distributions:

$$p(\mathbf{x};\theta) \quad = \quad \exp\Big\{ \sum_{\alpha} \theta_\alpha \phi_\alpha(\mathbf{x}) \; - \; \Phi(\theta) \Big\}$$

Log partition function:

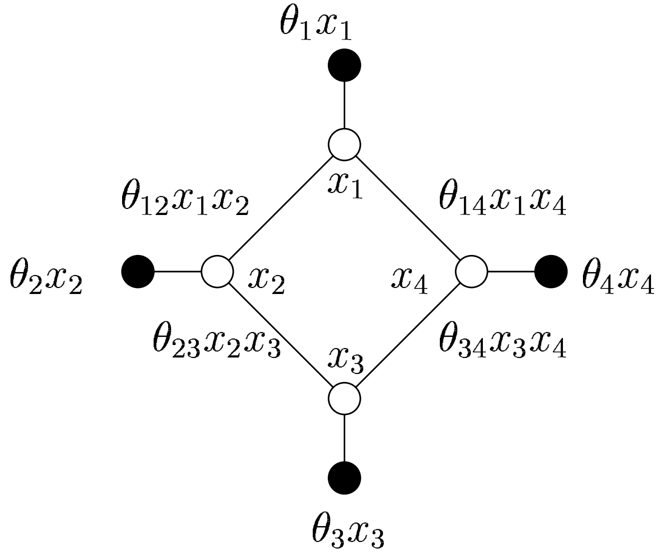
$$\Phi(\theta) \quad = \quad \log \Big(\sum_{\mathbf{x} \in \mathcal{X}^n} \exp\{ \sum_{\alpha} \theta_\alpha \phi_\alpha(\mathbf{x}) \} \Big)$$

$\boldsymbol{\phi} = \{ \phi_\alpha \mid \alpha \in \mathcal{I} \} \quad \equiv \quad$ collection of potential functions

$\theta = \{ \theta_\alpha \mid \alpha \in \mathcal{I} \} \quad \equiv \quad$ weights on potentials

Special case: Ising model

Binary variables on a graph with pairwise cliques



$$\begin{aligned} \phi &= \{\, x_s \mid s \in V \,\} \cup \{\, x_s x_t \mid (s,t) \in E \,\} \\ \mathcal{I} &= V \cup E \\ \mathcal{X}^n &= \{0,1\}^n \end{aligned}$$

$$p(\mathbf{x};\theta) = \exp\left\{ \sum_{s\in V} \theta_s x_s + \sum_{(s,t)\in E} \theta_{st} x_s x_t - \Phi(\theta) \right\}$$

An alternative view

Idea: Think about optimization not in terms of distributions p, but rather in terms of *only* the mean parameters:

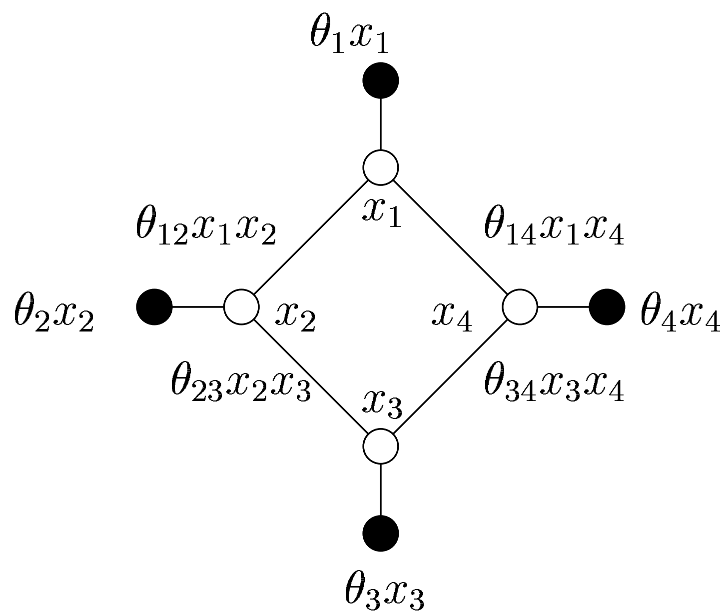
$$\mu_\alpha \quad := \quad \sum_{\mathbf{x}} p(\mathbf{x})\phi_\alpha(\mathbf{x})$$

Question: What is the relevant constraint set?

A marginal polytope is a set of realizable or globally consistent marginals:

$$\mathrm{MARG}(G; \boldsymbol{\phi}) \quad = \quad \Big\{\mu \in \mathbb{R}^d \;\Big|\; \mu = \sum_{\mathbf{x} \in \mathcal{X}^n} p(\mathbf{x})\boldsymbol{\phi}(\mathbf{x}) \qquad \text{for } \textit{some } p(\cdot)\Big\}$$

Ising model example

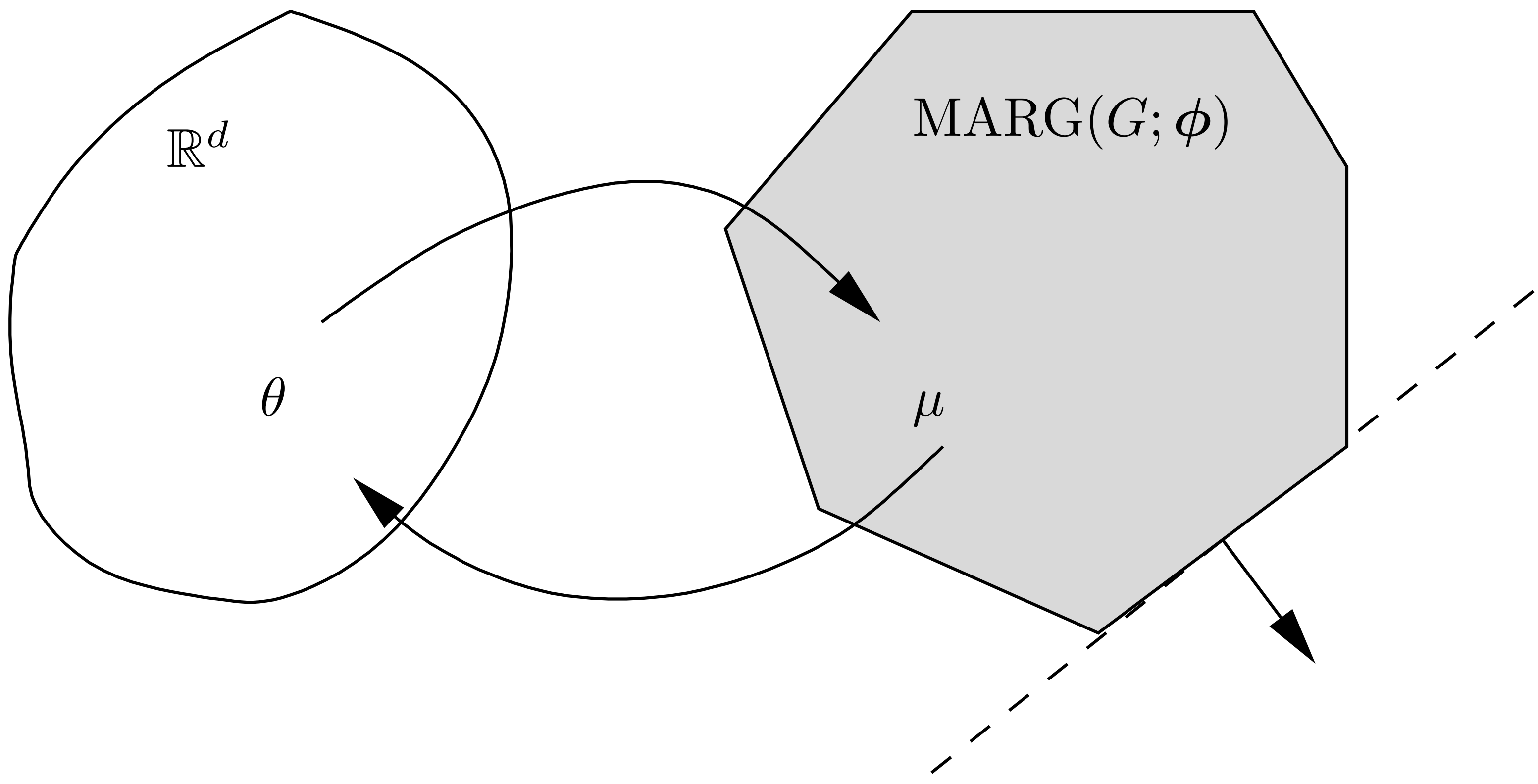


Potentials $\boldsymbol{\phi} = \{x_s \mid s \in V\} \cup \{x_s x_t \mid (s,t) \in E\}$

Relevant marginals $\mu_s = \mathbb{E}_\theta[x_s] \qquad \mu_{st} = \mathbb{E}_\theta[x_s x_t]$

Associated constraint set is known as the *correlation polytope* or the *binary quadric polytope.* (e.g., Deza & Laurent, 1997)

Geometry and moment mapping



Variational principle in terms of marginals

- the dual to $\Phi(\theta)$ has the form:

$$\Phi^*(\mu) \quad = \quad \begin{cases} -H(p(\mathbf{x};\theta(\mu))) & \text{if } \mu \in \mathrm{MARG}(G;\boldsymbol{\phi}) \\ +\infty & \text{otherwise.} \end{cases}$$

- leads to a representation of Φ in terms of Φ^*:

$$\underbrace{\Phi(\theta)}_{\text{log partition function}} \quad = \quad \underbrace{\max_{\mu \in \mathrm{MARG}(G;\boldsymbol{\phi})} \left\{ \langle \mu,\, \theta \rangle \; - \; \Phi^*(\mu) \right\}}_{\text{maximum entropy problem over marginal polytope}}$$

- moreover, maximum is attained uniquely at desired marginals:

$$\mu_\alpha \quad = \quad \sum_{\mathbf{x} \in \mathcal{X}^n} p(\mathbf{x};\theta)\phi_\alpha(\mathbf{x}) \; = \; \mathbb{E}_\theta[\phi_\alpha(\mathbf{x})].$$

Convex relaxations

Strategy: Obtain upper bounds by *relaxation* of original problem.

Requirements:

(a) convex outer approximation to marginal polytope $\mathrm{MARG}(G; \phi)$.

(b) concave upper bound on entropy function $-\Phi^*(\mu)$.

Tools:

(a) tree and hypertree approaches (Bethe/Kikuchi etc.)

(b) semidefinite methods

(c) combination of semidefinite and hypertree methods

Semidefinite outer bounds on marginal polytopes

Focus on:

(a) binary case with "spins" $\mathbf{x} \in \{-1, +1\}^n$.

(b) complete graph K_n on n nodes.

Refer to the associated marginal polytope as $\text{MARG}(K_n)$.

Relevant marginals:

$$\mu_s = \mathbb{E}_\theta[x_s] \quad \text{for all } s = 1, \ldots n$$

$$\mu_{st} = \mathbb{E}_\theta[x_s x_t] \quad \text{for all } (s, t)$$

Semidefinite outer bounds on binary marginal polytope.

(e.g., Laurent, 2001; Lasserre, 2001; Parrilo, 2002)

First order: Optimizing over covariance matrices

The covariance matrix of $\mathbf{x}$ must be positive semidefinite:

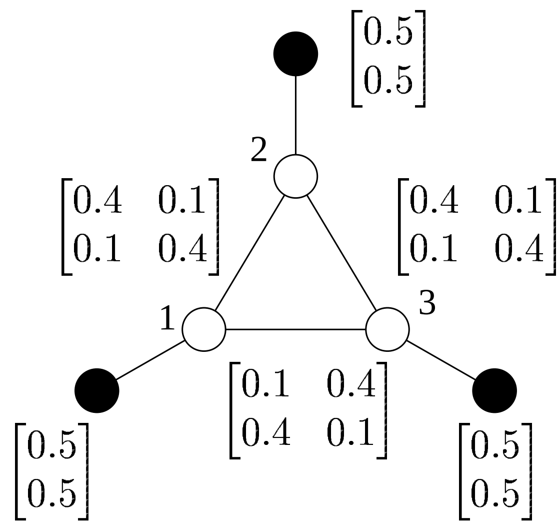
$$\mathrm{cov}(\mathbf{x}) \ = \ \mathbb{E}[\mathbf{x}\mathbf{x}^T] - \mathbb{E}[\mathbf{x}]\mathbb{E}[\mathbf{x}^T] \ \succeq \ 0$$

By Schur complement, equivalent to enforce PSD constraint on

$$\mathbb{E}\left\{ \begin{bmatrix} 1 \\ \mathbf{x} \end{bmatrix} \begin{bmatrix} 1 & \mathbf{x} \end{bmatrix} \right\} \ = \ \begin{bmatrix} 1 & | & \mu_1 & \mu_2 & \mu_3 & \cdots & \mu_n \\ -- & -- & -- & -- & -- & -- & -- \\ \mu_1 & | & 1 & \mu_{12} & \mu_{13} & \cdots & \mu_{1n} \\ \mu_2 & | & \mu_{21} & 1 & \mu_{23} & \cdots & \mu_{2n} \\ \mu_3 & | & \mu_{31} & \mu_{32} & 1 & \vdots & \mu_{3n} \\ \vdots & | & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mu_n & | & \mu_{n1} & \mu_{n2} & \mu_{n3} & \cdots & 1 \end{bmatrix}$$

Illustrative example

Tree-consistent (pseudo)marginals



Second-order moment matrix

$$\begin{bmatrix} \mu_1 & \mu_{12} & \mu_{13} \\ \mu_{21} & \mu_2 & \mu_{23} \\ \mu_{31} & \mu_{32} & \mu_3 \end{bmatrix} = \begin{bmatrix} 0.5 & 0.4 & 0.1 \\ 0.4 & 0.5 & 0.4 \\ 0.1 & 0.4 & 0.5 \end{bmatrix}$$

Not positive-semidefinite!

Concave upper bound on entropy

Challenge: Recall that entropy function $-\Phi^*(\mu)$ in terms of *only* μ lacks an explicit form.

For the Ising model, we have second-order information:

$$\mu_s := \mathbb{E}[x_s] \;\; \forall\, s \in V, \qquad \mu_{st} := \mathbb{E}[x_s x_t] \;\; \forall\, (s,t) \in E$$

Lemma: The differential entropy of any $\widetilde{\mathbf{x}}$ is upper-bounded by the covariance-matched Gaussian as follows:

$$h(\widetilde{\mathbf{x}}) \quad \leq \quad \frac{1}{2} \log \det \operatorname{cov}(\widetilde{\mathbf{x}}) + \frac{n}{2} \log(2\pi e)$$

Note: The *differential entropy* $h(\widetilde{\mathbf{x}}) := -\int p(\widetilde{\mathbf{x}}) \log p(\widetilde{\mathbf{x}}) d\widetilde{\mathbf{x}}$.

Log-determinant relaxation

Consider an outer bound $\text{OUT}(K_n)$ that satisfies:

$$\text{MARG}(K_n) \quad \subseteq \quad \text{OUT}(K_n) \; \subseteq \; \text{SDEF}_1(K_n)$$

Let $M_1(\mu) \in \text{OUT}(K_n)$ be a covariance matrix. Note that constraints imply that $M_1[\mu] \succeq 0$.

Log-det relaxation: For any such $\text{OUT}(K_n)$, $\Phi(\theta)$ is upper bounded by:

$$\max_{\mu \in \text{OUT}(K_n)} \left\{ \langle \theta, \, \mu \rangle + \frac{1}{2} \log \det \left[M_1(\mu) + \frac{1}{3} \, \text{blkdiag}[0, I_n] \right] \right\} + \frac{n}{2} \log(\frac{\pi e}{2})$$

Note: Such a log-det problem with LMI constraints can be solved efficiently by an interior-point method. (Vandenberghe, Boyd, & Wu, 1998)

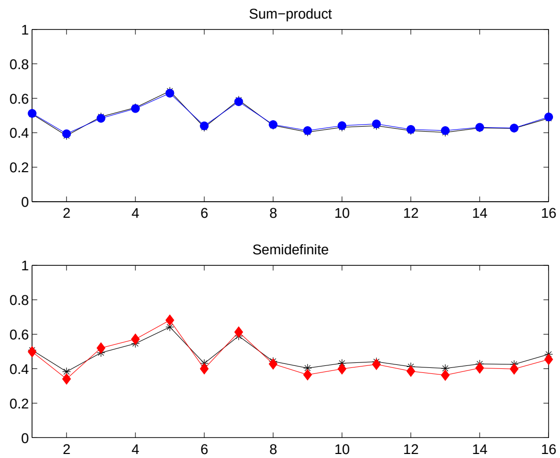
Results for fully connected graph

Problem type		Method			
		Sum-product		Log-determinant	
Coup.	Str.	Mean ± std	Range	Mean ± std	Range
−	Weak	0.037 ± 0.015	$[0.01, 0.10]$	0.020 ± 0.005	$[0.01, 0.03]$
−	Strong	0.071 ± 0.032	$[0.03, 0.20]$	0.018 ± 0.005	$[0.01, 0.04]$
+/−	Weak	0.004 ± 0.005	$[0.00, 0.04]$	0.020 ± 0.005	$[0.01, 0.03]$
+/−	Strong	0.055 ± 0.060	$[0.01, 0.31]$	0.021 ± 0.010	$[0.01, 0.06]$
+	Weak	0.024 ± 0.016	$[0.00, 0.08]$	0.027 ± 0.015	$[0.01, 0.06]$
+	Strong	0.435 ± 0.196	$[0.08, 0.86]$	0.033 ± 0.019	$[0.01, 0.09]$

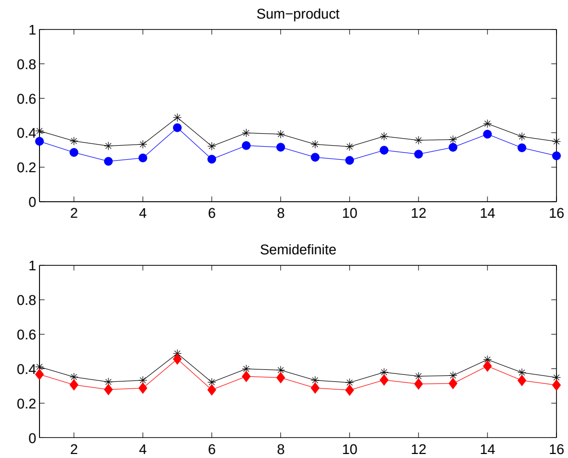
Results for nearest-neighbor grid

Problem type		Method			
		Sum-product		Log-determinant	
Coup.	Str.	Mean ± std	Range	Mean ± std	Range
$-$	Weak	0.294 ± 0.124	$[0.04, 0.59]$	0.047 ± 0.028	$[0.01, 0.12]$
$-$	Strong	0.342 ± 0.167	$[0.04, 0.78]$	0.041 ± 0.030	$[0.00, 0.12]$
$+/-$	Weak	0.014 ± 0.024	$[0.00, 0.20]$	0.016 ± 0.004	$[0.01, 0.02]$
$+/-$	Strong	0.095 ± 0.111	$[0.01, 0.54]$	0.038 ± 0.024	$[0.01, 0.11]$
$+$	Weak	0.440 ± 0.200	$[0.06, 0.90]$	0.047 ± 0.030	$[0.00, 0.13]$
$+$	Strong	0.520 ± 0.226	$[0.06, 0.94]$	0.042 ± 0.031	$[0.00, 0.12]$

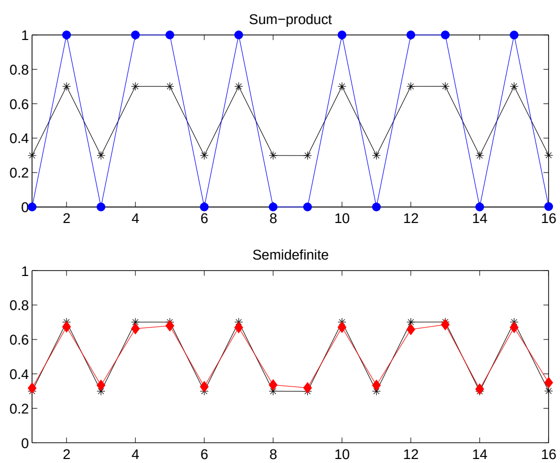
Sum-product versus log-determinant



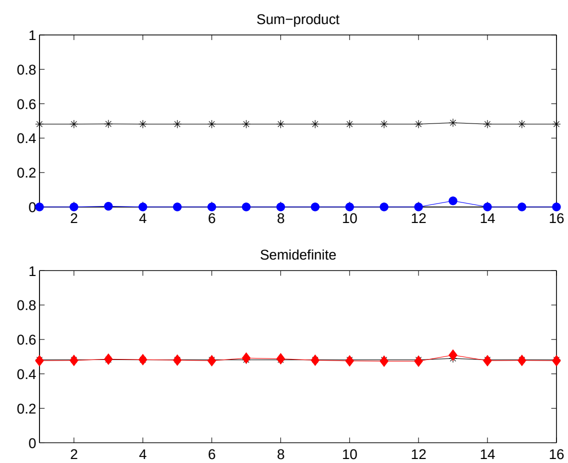
(a)



(b)



(c)



(d)

Summary

- role of mean parameters and marginal polytopes in variational principle

- log-determinant relaxation for approximate inference

- open questions:
  (a) relative roles of approximations to $\mathrm{MARG}(G)$ and entropy bound
  (b) performance guarantees for specific problem classes: link to integer programming results (e.g., Goemans & Williamson, 1995)
  (c) faster distributed techniques for solving relaxations

Contact information

Martin Wainwright

wainwrig@eecs.berkeley.edu

Papers at: http://www.eecs.berkeley.edu/~wainwrig

Supplementary material

Higher order extensions

1. Moment matrices involving higher-order multinomials.

   Example:

$$\text{cor}\left(1, x_1, x_2, x_1 x_2\right) \quad = \quad \begin{bmatrix} 1 & \mu_1 & \mu_2 & \mu_{12} \\ \mu_1 & 1 & \mu_{12} & \mu_2 \\ \mu_2 & \mu_{12} & 1 & \mu_1 \\ \mu_{12} & \mu_2 & \mu_1 & 1 \end{bmatrix} \succeq 0$$

2. For more general discrete spaces $\mathcal{X} = \{0, 1, \ldots, m-1\}$, consider correlations among vectors of monomials:

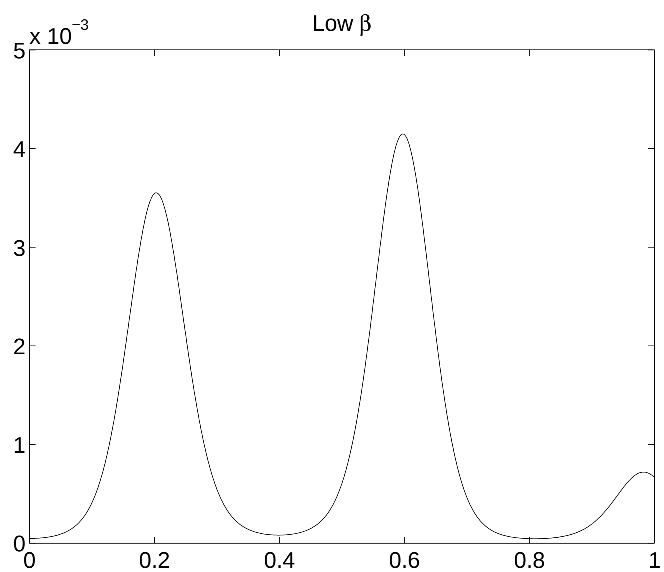
$$\mathcal{P}(s) \quad = \quad \{\, x_s,\, x_s^2, \ldots, x_s^{m-1}\}$$

Zero temperature limit

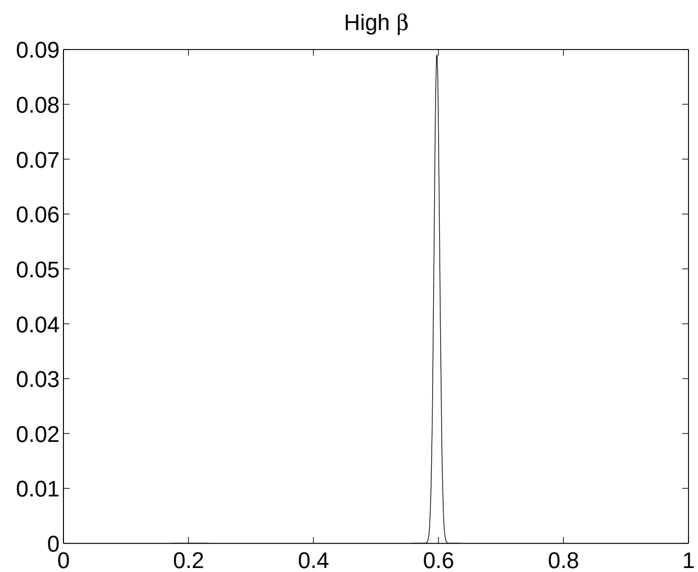
For fixed θ, consider the 1-parameter family of distributions:

$$p(\mathbf{x}; \beta\theta) \quad = \quad \exp\left\{\beta\langle\theta, \boldsymbol{\phi}(\mathbf{x})\rangle - \Phi(\beta\theta)\right\}$$

Here β should be viewed as inverse "temperature".



(a) Low β



(b) High β

Link to SDP relaxation for integer programming

For all $\beta > 0$, $\frac{1}{\beta}\Phi(\beta\theta)$ is upper bounded by the following:

$$\frac{1}{\beta} \max_{\mu \in \mathrm{OUT}(K_n)} \left\{ \langle \beta\theta,\, \mu \rangle + \frac{1}{2} \log\det \left[M_1(\mu) + \frac{1}{3}\, \mathrm{blkdiag}[0, I_n] \right] \right\} + C$$

Taking limits as $\beta \to \infty$ corresponds to computing a recession function.
(Rockafellar, 1970)

Result is a well-known SDP relaxation for integer programming:

$$\max_{\mathbf{x} \in \mathcal{X}^n} \langle \theta,\, \boldsymbol{\phi}(\mathbf{x}) \rangle \quad \leq \quad \max_{\mu \in \mathrm{OUT}(K_n)} \langle \theta,\, \mu \rangle$$

For strong coupling, behavior of log-det relaxation (for inference) approaches that of a SDP relaxation for integer programming.