Computing bounds for entropy of stationary $\mathbb{Z}^d$-Markov random fields

Brian Marcus (University of British Columbia)

Ronnie Pavlov (University of Denver)

WCI 2013
Hong Kong University
Dec. 11, 2013

Outline

- Stationary 1-D processes
  - Entropy (rate)
  - Markov chains
- Stationary 2-D processes
  - Entropy (rate)
  - 2-D Markov random fields (MRF)
  - Examples of MRFs: Gibbs measures (e.g., hardcore, Ising, Potts)
- Exponential Strong spatial mixing (ESSM)
- Main result: For any fixed 2-D stationary Gibbs measure, satisfying ESSM, there is an efficient algorithm to approximate entropy (but *only* in dimension 2).
- Others get efficient algorithms for higher dimensional stationary Gibbs measures, by establishing ESSM and computation tree for special processes.
- Proof sketch

Entropy of 1-D stationary processes

Entropy (rate) of a 1-D stationary process $X = \ldots X_{-1}, X_0, X_1, \ldots$ (discrete time, finite-valued) is defined:

$$h(X) = \lim_{n\to\infty} \frac{H(X_1 \ldots X_n)}{n} = \lim_{n\to\infty} H(X_0 | X_{-n}, \ldots, X_{-1})$$

Entropy of a stationary 1-D Markov chain, with transition matrix P and stationary vector π, has a closed form expression:

$$h(X) = -\sum_{ij} \pi_i P_{ij} \log P_{ij}$$

Example:

$$P = \begin{bmatrix} \frac{1}{\lambda} & \frac{1}{\lambda^2} \\ 1 & 0 \end{bmatrix}$$

where $\lambda = \frac{1+\sqrt{5}}{2}$.

$$h(X) = \log \lambda \approx .69.$$

Note: X is supported on the Golden Mean constraint: forbid adjacent 1's.

Fact: this process is uniform in the sense that, for fixed values a, b,

$$\text{Prob}(x_1 \ldots x_n | x_0 = a, x_{n+1} = b)$$

is the uniform distribution on *allowed* sequences $a x_1 \ldots x_n b$.

Entropy of 2-D stationary processes

Entropy (rate) of a 2-D stationary process $X = X_{ij}$, is defined

$$h(X) = \lim_{n\to\infty} \frac{H(X_{i,j})_{1\le i,j\le n}}{n^2}$$

View process as a translation-invariant measure μ on $\mathcal{A}^{\mathbb{Z}^2}$ (set of configurations on the 2-dimensional integer lattice with finite alphabet $\mathcal{A}$).

Markov random field:

A measure μ on $\mathcal{A}^{\mathbb{Z}^2}$ such that for any

- finite $S \subset \mathbb{Z}^2$,
- configuration x on S ($x \in \mathcal{A}^S$)
- finite $T \subset \mathbb{Z}^2$ s.t. $\partial S \subseteq T \subseteq \mathbb{Z}^2 \setminus S$
- configuration δ on T with $\mu(\delta) > 0$,

$$\mu(x \mid \delta) = \mu(x \mid \delta|_{\partial S})$$

uniform MRF: for a given hard constraint, each $\mu(\cdot \mid \delta|_{\partial S})$ is uniform.

Example: uniform 2-D hard square measure: forbid adjacent 1's, horizontally and vertically.

$$h(X) \approx .59.$$

Much harder to compute entropy of 2-D stationary MRF, even in the uniform case. Known only in a handful of special cases.

Q: What would be a satisfactory "formula" for the entropy?

One possible answer: an **efficient algorithm**: on input ϵ produces numbers h^+, h^- s.t.

- $h^- \leq h(X) \leq h^+$
- $h^+ - h^- < \epsilon$
- h^+, h^- are computed in time poly$(1/\epsilon)$.

Fact: entropy of uniform 2-D Hard square measure has an efficient algorithm (Weitz, Gamarnik-Katz; independently, Pavlov)

Notation

Let μ be a stationary 2-D process.

Given finite set $S \subset \mathbb{Z}^d$,

$$H_\mu(S) := \sum_{w \in \mathcal{A}^S} -\mu(w) \log(\mu(w))$$

In this notation Entropy of μ is:

$$h(\mu) = \lim_{n \to \infty} \frac{H_\mu(B_n)}{n^2},$$

where B_n is an $n \times n$ square.

Conditional entropy: for finite disjoint S, T,

$$H_\mu(S \mid T) := \sum_{w \in \mathcal{A}^S, y \in \mathcal{A}^T: \ \mu(y) > 0} -\mu(w, y) \log \mu(w \mid y)$$

Extend to infinite T:

$$H_\mu(S \mid T) := \lim_n H_\mu(S \mid T_n)$$

for a nested sequence of finite sets $T_1 \subset T_2 \subset \ldots$ with $\cup_n T_n = T$

Conditional entropy formula:

$$h(\mu) = H_\mu(\mathbf{0} \mid \mathcal{P}^-).$$

where $\mathcal{P}^- = \{z \in \mathbb{Z}^2 : \ z \preceq \mathbf{0}\}$, the lexicographic past of the origin.

Exponential Strong Spatial Mixing (ESSM)

A stationary 2-D MRF μ satisfies **exponential strong spatial mixing** (ESSM) if there exist $C, \alpha > 0$ s.t. s.t

- for any disjoint S, T, with $S, T \subset B_n$
- any $x \in \mathcal{A}^S$, $y \in \mathcal{A}^T$, $\delta, \delta' \in \mathcal{A}^{\partial B_n}$ s.t. $\mu(y, \delta), \mu(y, \delta') > 0$,

$$|\mu(x \mid y, \delta) - \mu(x \mid y, \delta')| < C|S|e^{-\alpha d(S, \partial B_n)}$$

Fact: 2-D uniform hard squares satisfies ESSM (Weitz, Gamarnik-Katz; independently, Pavlov)

Exponentially tight upper and lower bounds

Let $\mathcal{P}^+ = \{z \in \mathbb{Z}^2 : \ z \succeq \mathbf{0}\}$, the lexicographic future of the origin. Let $S_n = B_n \cap \mathcal{P}^+$, and $U_n = B_n \cap \partial\mathcal{P}^+$.

Lemma 1: For a stationary 2-D MRF μ,

$$H_\mu(\mathbf{0} \mid \partial S_n) \le h(\mu) \le H_\mu(\mathbf{0} \mid U_n).$$

Proof:

$$H_\mu(\mathbf{0} \mid \partial S_n) = H_\mu(\mathbf{0} \mid \partial S_n, \mathcal{P}^-) \le H_\mu(\mathbf{0} \mid \mathcal{P}^-) = h(\mu) \le H_\mu(\mathbf{0} \mid U_n).$$

Theorem 1: For a stationary 2-D MRF μ that satisfies ESSM

$$|H_\mu(\mathbf{0} \mid U_n) - H_\mu(\mathbf{0} \mid \partial S_n)| \ = \ O(n)e^{-\alpha n}$$

Proof idea: For $y \in \mathcal{A}^{U_n}$, let

$$E(y) = \{w \in \mathcal{A}^{\partial S_n \setminus U_n} : \mu(wy) > 0\}$$

Then

$$H_\mu(\mathbf{0} \mid U_n) = \sum_{y \in \mathcal{A}^{U_n}:\ \mu(y)>0} \mu(y) H_\mu(\mathbf{0} \mid y)$$

and

$$H_\mu(\mathbf{0} \mid \partial S_n) = \sum_{y \in \mathcal{A}^{U_n}:\ \mu(y)>0} \mu(y) \sum_{w \in E(y)} \mu(w \mid y) H_\mu(\mathbf{0} \mid yw).$$

By ESSM, for all $x \in \mathcal{A}^{\mathbf{0}}$, $y \in \mathcal{A}^{U_n}$, and $w, w' \in E(y)$,

$$|\mu(x \mid yw) - \mu(x \mid yw')| \le Ce^{-\alpha n}.$$

And

$$\mu(x \mid y) = \sum_{w \in E(y)} \mu(w \mid y) \mu(x \mid yw),$$

Apply Jensen's inequality.

Note: Lemma 1 and Theorem 1 extend to any dimension.

Q: How to efficiently approximate these upper and lower bounds?

Stationary 2-D Gibbs measures:

MRFs are not so tractable in general: too many conditional probability measures.

nearest-neighbour interaction: function Φ, on configurations on single sites and adjacent sites. Let

For finite set $S \subset \mathbb{Z}^2$, and any $w \in \mathcal{A}^S$, the **energy function** is defined

$$U_S^\Phi(w) := \sum_e \Phi(w(e)) + \sum_v \Phi(w(v))$$

where the sums range over all edges e and vertices v of S.

The **partition function** of $\Phi, S, \delta \in \mathcal{A}^{\partial S}$ is

$$Z^{\Phi,\delta}(S) := \sum_{w \in \mathcal{A}^S} e^{-U_S^\Phi(w\delta)}.$$

For any nearest-neighbor interaction Φ, an MRF μ is called a **Gibbs measure** for Φ if for any finite set $S \subset \mathbb{Z}^2$ and $\delta \in \mathcal{A}^{\partial S}$ for which $\mu(\delta) > 0$, we have $Z^{\Phi,\delta}(S) \neq 0$ and, for any $w \in \mathcal{A}^S$,

$$\mu(w|\delta) = \frac{e^{-U_S^\Phi(w\delta)}}{Z^{\Phi,\delta}(S)}.$$

Classical Gibbs measures

- Ising model: $\mathcal{A} = \{\pm 1\}$, $\Phi(a) = -\beta E a$, $\Phi(a,b) = -\beta J ab$ for constants E (external magnetic field), J (coupling strength), and β (inverse temperature).

- n-state Potts model: $\mathcal{A} = \{1, \dots, n\}$, $\Phi(a) = 0$, $\Phi(a,b) = -\beta J$ if $a = b$ and 0 otherwise.

- uniform n-coloring measure: $\mathcal{A} = \{1, \dots n\}$, $\Phi(a) = 0$, $\Phi(a,b) = \infty$ if $a = b$ and 0 otherwise; this can be thought of as the limiting case of the n-state Potts model as $\beta \to -\infty$.

- uniform hard square measure: $\mathcal{A} = \{0, 1\}$, $\Phi(a) = 0$, $\Phi(a,b) = 0$ unless $a = b = 1$ in which case $\Phi(a,b) = \infty$

Efficient approximation of the upper and lower bounds.

Theorem 2: Let μ be a stationary 2-D Gibbs measure which satisfies ESSM. Let K_n satisfy $K_n \subset B_n$ and $|K_n| = O(n)$. Then for some for some $C', \alpha' > 0$. there is an algorithm which, on input n, computes upper and lower bounds to $H_\mu(\mathbf{0} \mid K_n)$ in time $e^{O(n)}$ to within tolerance $C'e^{-\alpha' n}$.

Corollary (Main Result): Let μ be a stationary 2-D Gibbs measure which satisfies ESSM. Then there is an algorithm which, on input $\epsilon > 0$, computes upper and lower bounds to $h(\mu)$ in time poly$(1/\epsilon)$ within tolerance ϵ.

Note; algorithm does *not* require knowledge of decay rate α of ESSM.

Proof of Corollary: Write $\epsilon = Cne^{-\alpha n} + C'e^{-\alpha' n}$. By Lemma 1 and Theorem 1, $H_\mu(\mathbf{0} \mid \partial S_n)$ and $H_\mu(\mathbf{0} \mid U_n)$ are upper and lower approximations to within tolerance ϵ. Apply Theorem 2 to $K_n = \partial S_n$ and $K_n = U_n$. Each can be approximated to within tolerance ϵ in time $e^{O(n)} = \text{poly}(1/\epsilon)$.

Note: Proof extends to any dimension to give an algorithm, but is not efficient.

Sketch of Proof of Theorem 2:

Fix $L > 0$.

Step 1: For $w \in \mathcal{A}^{K_n}$, and $\delta \in \mathcal{A}^{\partial B_{n+Ln}}$, compute $\mu(w|\delta)$ exactly.

Idea: Set up modified transfer matrix to compute $U_S^\Phi(w)$ and $Z_S^{\Phi,\delta}(B_{n+Ln})$ exactly.

computation error $= 0$.

computation time $= e^{O(n+Ln)}$.

Step 2: For $w \in \mathcal{A}^{K_n}$ find $\delta_w^\pm \in \mathcal{A}^{\partial B_{n+Ln}}$ s.t.

$$\mu^-(w) := \mu(w|\delta_w^-) \le \mu(w) \le \mu(w|\delta_w^+) =: \mu^+(w)$$

Idea: $\mu(w) = \sum_\delta \mu(w|\delta)\mu(\delta)$.

computation error $= e^{-\alpha L n}$ (by ESSM)

computation time $= e^{O(n+Ln)}$ (brute force)

Step 3: Do same as in Step 2 for $\mu(x_0|w)$ instead of $\mu(w)$

Step 4: Form $H_\mu^\pm(\mathbf{0} \mid K_n)$ by replacing each $\mu(w)$ by $\mu^\pm(w)$ and each $\mu(x_0|w)$ by $\mu^\pm(x_0|w)$.

computation error $= e^{O(n)}e^{-\alpha L n}$

computation time $= e^{O(n)}e^{O(n+Ln)} = e^{O(n)}$

If L is sufficiently large (depending on $|\mathcal{A}|, \alpha$ and some constants)

computation error $= C'e^{-\alpha' n}$ for some $C', \alpha' > 0$.