KNOW Center

AUSTRIA'S LEADING RESEARCH CENTER
FOR DATA-DRIVEN BUSINESS AND BIG DATA ANALYTICS

On Functions of Markov Random Fields

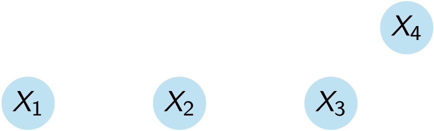
IEEE Information Theory Workshop 2020

Bernhard C. Geiger & Ali Al-Bashabsheh

The Authors and Funders

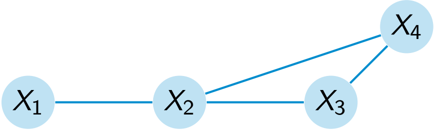
H2020 No. 824115

Functions of Markov Random Fields (MRFs)



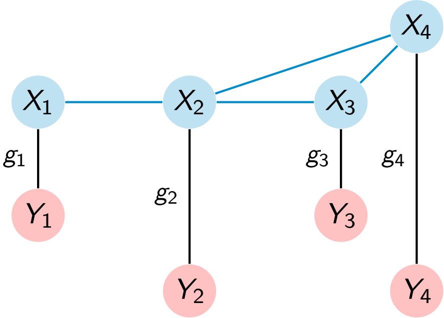
- $\mathcal{V} = \{1, 2, 3, 4\}$
- set of RVs X_i
- alphabet $\mathcal{X}_i$
- PMF p_{X_i}
- $X = (X_1, \dots, X_4)$

Functions of Markov Random Fields (MRFs)



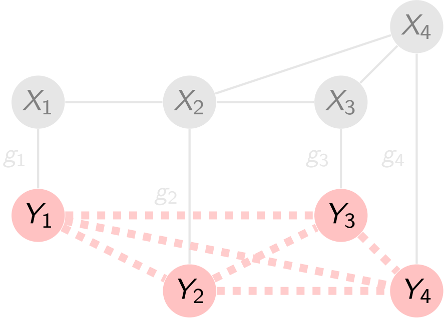
- $\mathcal{G} = (\mathcal{V}, E)$
- $\mathcal{N}_i$ are neighbors of i
- X is a $(\mathcal{G}, p_X)$-MRF
- $p_{X_i|X_{i/}} = p_{X_i|X_{\mathcal{N}_i}}$
- e.g., $p_{X_3|X_1,X_2,X_4} = p_{X_3|X_2,X_4}$

Functions of Markov Random Fields (MRFs)



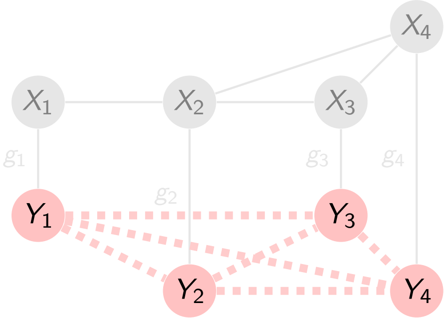
- $g_i\colon \mathcal{X}_i \to \mathcal{Y}_i$
- g_i non-injective
- $Y_i = g_i(X_i)$
- $Y = (Y_1, \ldots, Y_4)$

Functions of Markov Random Fields (MRFs)



- W.r.t. which graph $\mathcal{G}_Y = (\mathcal{V}, E_Y)$ is Y an MRF?
- **Lumpability:** $E_Y \subseteq E$

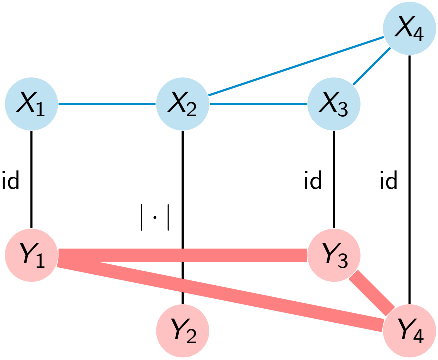
Functions of Markov Random Fields (MRFs)



- W.r.t. which graph $\mathcal{G}_Y = (\mathcal{V}, E_Y)$ is Y an MRF?
- **Lumpability:** $E_Y \subseteq E$

$$p_{X_i|X_{\not i}} = p_{X_i|X_{\mathcal{N}_i}} \Leftrightarrow H(X_i|X_{\not i}) = H(X_i|X_{\mathcal{N}_i}) \text{ but } H(Y_i|Y_{\not i}) \leq H(Y_i|Y_{\mathcal{N}_i})$$

Running Example



- $\mathcal{X}_i = \{-1, 1\}$
- not lumpable

Related Work

- Lumpability of Markov chains ($\mathcal{G}$ is a directed line graph)[1]
  - linear algebraic conditions[2]
  - information-theoretic conditions[3]
- Yeung et al.[4,5] did not specify p_X and considered sub-graphs, i.e., g_i is either constant or identity
- Perez & Heitz[6] specified p_X and considered sub-graphs or stochastic maps $p_{Y_i|X_i} > 0$
- **This work:** p_X specified, general g_i

---

[1] Kemeny and Snell, *Finite Markov Chains*, 1976

[2] Gurvits and Ledoux, "Markov property for a function of a Markov chain: a linear algebra approach", 2005

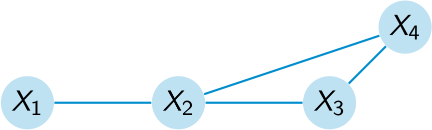
[3] Geiger and Temmel, "Lumpings of Markov chains, entropy rate preservation, and higher-order lumpability", 2014

[4] Yeung et al., "Information-theoretic characterizations of Markov random fields and subfields", 2017

[5] Yeung et al., "On Information-Theoretic Characterizations of Markov Random Fields and Subfields", 2019

[6] Perez and Heitz, "Restriction of a Markov random field on a graph and multiresolution statistical image modeling", 1996

MRFs with positive p_X

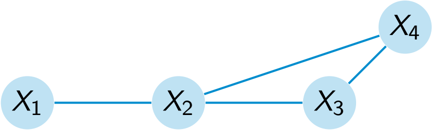


Lemma (Hammersley-Clifford)

X is a $(\mathcal{G}, p_X)$-MRF iff there exists a family of clique potential functions $\{\psi_C,\ C \in \mathcal{C}\}$ such that

$$\forall x \in \mathcal{X}: \quad p_X(x) = \frac{1}{Z} \prod_{C \in \mathcal{C}} \psi_C(x_C),$$

MRFs with positive p_X



Lemma (Hammersley-Clifford)

X is a $(\mathcal{G}, p_X)$-MRF iff there exists a family of clique potential functions $\{\psi_C,\ C \in \mathcal{C}\}$ such that

$$\forall x \in \mathcal{X}: \quad p_X(x) = \frac{1}{Z} \prod_{C \in \mathcal{C}} \psi_C(x_C),$$

E.g., $p_X(x) \propto \psi_{\{1,2\}}(x_1, x_2) \cdot \psi_{\{2\}}(x_2) \cdot \psi_{\{2,3,4\}}(x_2, x_3, x_4)$

A Trivial Sufficient Condition

- Y is a $(\mathcal{G}, p_Y)$-MRF if

$$p_Y(y) = \frac{1}{Z'} \prod_{C \in \mathcal{C}} U_C(y_C)$$

A Trivial Sufficient Condition

- Y is a $(\mathcal{G}, p_Y)$-MRF if

$$p_Y(y) = \frac{1}{Z'} \prod_{C \in \mathcal{C}} U_C(y_C)$$

- Since $Y = g(X)$

$$p_Y(y) = \sum_{x \in g^{-1}(y)} p_X(x)$$

A Trivial Sufficient Condition

- Y is a $(\mathcal{G}, p_Y)$-MRF if

$$p_Y(y) = \frac{1}{Z'} \prod_{C \in \mathcal{C}} U_C(y_C)$$

- Since $Y = g(X)$ and X is a $(\mathcal{G}, p_X)$-MRF, we have

$$p_Y(y) = \sum_{x \in g^{-1}(y)} p_X(x) = \sum_{x \in g^{-1}(y)} \frac{1}{Z} \prod_{C \in \mathcal{C}} \psi_C(x_C)$$

A Trivial Sufficient Condition

- Y is a $(\mathcal{G}, p_Y)$-MRF if

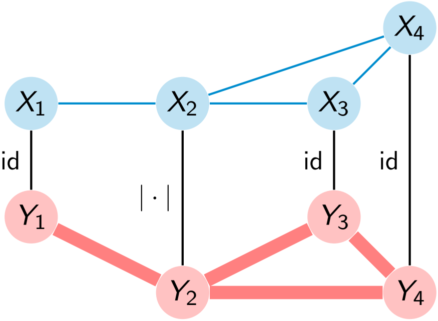
$$p_Y(y) = \frac{1}{Z'} \prod_{C \in \mathcal{C}} U_C(y_C)$$

- Since $Y = g(X)$ and X is a $(\mathcal{G}, p_X)$-MRF, we have

$$p_Y(y) = \sum_{x \in g^{-1}(y)} p_X(x) = \sum_{x \in g^{-1}(y)} \frac{1}{Z} \prod_{C \in \mathcal{C}} \psi_C(x_C)$$

- Y is a $(\mathcal{G}, p_Y)$-MRF if ψ_C are constant on the preimages under g

A Trivial Sufficient Condition



- $\mathcal{X}_i = \{-1, 1\}$
- p_X characterized by $\psi_{\{1,2\}}$, $\psi_{\{2\}}$,$\psi_{\{2,3,4\}}$
- lumpable if:

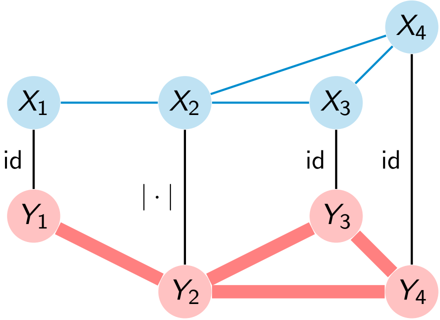
$$\psi_{\{2\}}(-1) = \psi_{\{2\}}(1)$$
AND
$$\psi_{\{1,2\}}(x_1, -1) = \psi_{\{1,2\}}(x_1, 1)$$
AND
$$\psi_{\{2,3,4\}}(-1, x_3, x_4) = \psi_{\{2,3,4\}}(1, x_3, x_4)$$

A Slightly Less Trivial Sufficient Condition

Theorem (Loosely)

Y is a MRF w.r.t. the graph $\mathcal{G}$ if for every vertex $i \in \mathcal{V}$ there is at most one clique potential ψ_C that is **not** *constant on the preimage under g*

A Slightly Less Trivial Sufficient Condition



- $\mathcal{X}_i = \{-1, 1\}$
- p_X characterized by $\psi_{\{1,2\}}$, $\psi_{\{2\}}$, $\psi_{\{2,3,4\}}$
- lumpable if:

$$\psi_{\{1,2\}}(x_1, -1) = \psi_{\{1,2\}}(x_1, 1) \text{ AND } \psi_{\{2\}}(-1) = \psi_{\{2\}}(1)$$

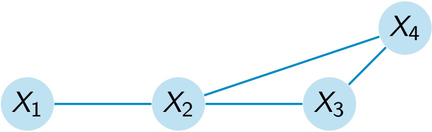
OR

$$\psi_{\{2,3,4\}}(-1, x_3, x_4) = \psi_{\{2,3,4\}}(1, x_3, x_4) \text{ AND } \psi_{\{2\}}(-1) = \psi_{\{2\}}(1)$$

OR

$$\dots$$

An Information-Theoretic Sufficient Condition



Theorem

Y is a $(\mathcal{G}, p_Y)$-MRF if, for every $i \in \mathcal{V}$,

$$H(Y_i | Y_{\mathcal{N}_i}) = H(Y_i | X_{\mathcal{N}_i})$$

An Information-Theoretic Sufficient Condition

Theorem

Y is a $(\mathcal{G}, p_Y)$-MRF if, for every $i \in \mathcal{V}$,

$$H(Y_i | Y_{\mathcal{N}_i}) = H(Y_i | X_{\mathcal{N}_i})$$

Proof:

$$H(Y_i | Y_{\mathcal{N}_i}) \geq H(Y_i | Y_{\not i}) \geq H(Y_i | X_{\not i}) = H(Y_i | X_{\mathcal{N}_i})$$

An Information-Theoretic Sufficient Condition

Theorem (Sufficient Condition for Markov Chain Lumpability[7])

Let $X = (X_1, X_2, \dots)$ be a stationary Markov chain and let $Y_i = g_0(X_i)$. Then, $Y = (Y_1, Y_2, \dots)$ is a stationary Markov chain if, for some i,

$$H(Y_i|Y_{i-1}) = H(Y_i|X_{i-1})$$

- $H(Y_i|Y_{\mathcal{N}_i}) = H(Y_i|X_{\mathcal{N}_i})$ vs. $H(Y_i|Y_{i-1}) = H(Y_i|X_{i-1})$

[7] Geiger and Temmel, "Lumpings of Markov chains, entropy rate preservation, and higher-order lumpability", 2014

An Information-Theoretic Sufficient Condition

Theorem (Sufficient Condition for Markov Chain Lumpability[7])

Let $X = (X_1, X_2, \dots)$ be a stationary Markov chain and let $Y_i = g_0(X_i)$. Then, $Y = (Y_1, Y_2, \dots)$ is a stationary Markov chain if, for some i,

$$H(Y_i|Y_{i-1}) = H(Y_i|X_{i-1})$$

- $H(Y_i|Y_{\mathcal{N}_i}) = H(Y_i|X_{\mathcal{N}_i})$ vs. $H(Y_i|Y_{i-1}) = H(Y_i|X_{i-1})$
- $X_1 - X_2 - X_3 - \cdots$ vs. $X_1 \to X_2 \to X_3 \to \cdots$

[7] Geiger and Temmel, "Lumpings of Markov chains, entropy rate preservation, and higher-order lumpability", 2014

An Information-Theoretic Sufficient Condition

Theorem (Sufficient Condition for Markov Chain Lumpability[7])

Let $X = (X_1, X_2, \dots)$ be a stationary Markov chain and let $Y_i = g_0(X_i)$. Then, $Y = (Y_1, Y_2, \dots)$ is a stationary Markov chain if, for some i,

$$H(Y_i|Y_{i-1}) = H(Y_i|X_{i-1})$$

- $H(Y_i|Y_{\mathcal{N}_i}) = H(Y_i|X_{\mathcal{N}_i})$ vs. $H(Y_i|Y_{i-1}) = H(Y_i|X_{i-1})$
- $X_1 - X_2 - X_3 - \cdots$ vs. $X_1 \to X_2 \to X_3 \to \cdots$
- Undirected vs. directed graph $\mathcal{G}$

[7] Geiger and Temmel, "Lumpings of Markov chains, entropy rate preservation, and higher-order lumpability", 2014

An Information-Theoretic Sufficient Condition

Theorem (Sufficient Condition for Markov Chain Lumpability[7])

Let $X = (X_1, X_2, \dots)$ be a stationary Markov chain and let $Y_i = g_0(X_i)$. Then, $Y = (Y_1, Y_2, \dots)$ is a stationary Markov chain if, for some i,

$$H(Y_i|Y_{i-1}) = H(Y_i|X_{i-1})$$

- $H(Y_i|Y_{\mathcal{N}_i}) = H(Y_i|X_{\mathcal{N}_i})$ vs. $H(Y_i|Y_{i-1}) = H(Y_i|X_{i-1})$
- $X_1 - X_2 - X_3 - \cdots$ vs. $X_1 \to X_2 \to X_3 \to \cdots$
- Undirected vs. directed graph $\mathcal{G}$
- Suggests further work for Bayesian networks

[7] Geiger and Temmel, "Lumpings of Markov chains, entropy rate preservation, and higher-order lumpability", 2014

Conclusion

- **When is a function of an MRF an MRF on a subgraph?**
- **Two sufficient conditions:**
  - via clique potentials of equivalent Gibbs field
  - information-theoretic condition
- **Further results:**
  - conditions for Y to have the same entropy as X
  - information preservation, lossless compression

Conclusion

- **When is a function of an MRF an MRF on a subgraph?**
- **Two sufficient conditions:**
  - via clique potentials of equivalent Gibbs field
  - information-theoretic condition
- **Further results:**
  - conditions for Y to have the same entropy as X
  - information preservation, lossless compression

Thanks!