Information Theory and Image/Video Coding

Ming Jiang

School of Mathematical Sciences
Peking University

ming-jiang@pku.edu.cn

March 12, 2012

Outline

Images as Random Fields

- A monochrome digital image is presented as a matrix with pixel values corresponding to the intensity of light.
- Each pixel value is modeled as a `random variable`.
  - Image attributes are rarely deterministic;
  - They are generally characterized by correlations and likelihoods.
- Images are `random fields`.

- References
  - i [Geman and Geman, 1984].
  - ii [Geman and Graffigne, 1987].
  - iii [Geman, 1990].

Images as Random Fields

- A monochrome digital image is presented as a matrix with pixel values corresponding to the intensity of light.
- Each pixel value is modeled as a `random variable`.
  - Image attributes are rarely deterministic;
  - They are generally characterized by correlations and likelihoods.
- Images are `random fields`.

- References
  i [Geman and Geman, 1984].
  ii [Geman and Graffigne, 1987].
  iii [Geman, 1990].

Images as Random Fields

- A monochrome digital image is presented as a matrix with pixel values corresponding to the intensity of light.
- Each pixel value is modeled as a `random variable`.
  - Image attributes are rarely deterministic;
  - They are generally characterized by correlations and likelihoods.
- Images are `random fields`.

- References
  - i [Geman and Geman, 1984].
  - ii [Geman and Graffigne, 1987].
  - iii [Geman, 1990].

Images as Random Fields

- A monochrome digital image is presented as a matrix with pixel values corresponding to the intensity of light.
- Each pixel value is modeled as a `random variable`.
  - Image attributes are rarely deterministic;
  - They are generally characterized by correlations and likelihoods.
- Images are `random fields`.

- References
  i [Geman and Geman, 1984].
  ii [Geman and Graffigne, 1987].
  iii [Geman, 1990].

Images as Random Fields

- A monochrome digital image is presented as a matrix with pixel values corresponding to the intensity of light.
- Each pixel value is modeled as a `random variable`.
  - Image attributes are rarely deterministic;
  - They are generally characterized by correlations and likelihoods.
- Images are `random fields`.

- References
  - i [Geman and Geman, 1984].
  - ii [Geman and Graffigne, 1987].
  - iii [Geman, 1990].

Images as Random Fields

- A monochrome digital image is presented as a matrix with pixel values corresponding to the intensity of light.
- Each pixel value is modeled as a `random variable`.
  - Image attributes are rarely deterministic;
  - They are generally characterized by correlations and likelihoods.
- Images are `random fields`.

- References
  i [Geman and Geman, 1984].
  ii [Geman and Graffigne, 1987].
  iii [Geman, 1990].

Images as Random Fields

- A monochrome digital image is presented as a matrix with pixel values corresponding to the intensity of light.
- Each pixel value is modeled as a `random variable`.
  - Image attributes are rarely deterministic;
  - They are generally characterized by correlations and likelihoods.
- Images are `random fields`.

- References
  i [Geman and Geman, 1984].
  ii [Geman and Graffigne, 1987].
  iii [Geman, 1990].

Images as Random Fields

- A monochrome digital image is presented as a matrix with pixel values corresponding to the intensity of light.
- Each pixel value is modeled as a `random variable`.
  - Image attributes are rarely deterministic;
  - They are generally characterized by correlations and likelihoods.
- Images are `random fields`.
- References
  - i [Geman and Geman, 1984].
  - ii [Geman and Graffigne, 1987].
  - iii [Geman, 1990].

Images as Random Fields

- A monochrome digital image is presented as a matrix with pixel values corresponding to the intensity of light.
- Each pixel value is modeled as a `random variable`.
  - Image attributes are rarely deterministic;
  - They are generally characterized by correlations and likelihoods.
- Images are `random fields`.

- References
  i [Geman and Geman, 1984].
  ii [Geman and Graffigne, 1987].
  iii [Geman, 1990].

Spatial Dependence & Markov Random Fields

- Each pixel value
  - depends only on neighboring pixel values;
  - is independent from far away pixel values.
- Markov random fields provide a flexible mechanism for
  - modeling spatial dependence,
  - image attributes.
- General references for this section are
  i [Geman, 1990].
  ii [Wrinkler, 1995].
  iii [Li, 2009].
- Most of the materials is copy-edited from [Geman, 1990].

Spatial Dependence & Markov Random Fields

- Each pixel value
  - depends only on neighboring pixel values;
  - is independent from far away pixel values.
- Markov random fields provide a flexible mechanism for
  - modeling spatial dependence,
  - image attributes.
- General references for this section are
  i [Geman, 1990].
  ii [Wrinkler, 1995].
  iii [Li, 2009].
- Most of the materials is copy-edited from [Geman, 1990].

Spatial Dependence & Markov Random Fields

- Each pixel value
  - depends only on neighboring pixel values;
  - is independent from far away pixel values.
- Markov random fields provide a flexible mechanism for
  - modeling spatial dependence,
  - image attributes.
- General references for this section are
  - i [Geman, 1990].
  - ii [Wrinkler, 1995].
  - iii [Li, 2009].
- Most of the materials is copy-edited from [Geman, 1990].

Spatial Dependence & Markov Random Fields

- Each pixel value
  - depends only on neighboring pixel values;
  - is independent from far away pixel values.
- Markov random fields provide a flexible mechanism for
  - modeling spatial dependence,
  - image attributes.
- General references for this section are
  i [Geman, 1990].
  ii [Wrinkler, 1995].
  iii [Li, 2009].
- Most of the materials is copy-edited from [Geman, 1990].

Spatial Dependence & Markov Random Fields

- Each pixel value
  - depends only on neighboring pixel values;
  - is independent from far away pixel values.
- Markov random fields provide a flexible mechanism for
  - modeling spatial dependence,
  - image attributes.
- General references for this section are
  i. [Geman, 1990].
  ii. [Wrinkler, 1995].
  iii. [Li, 2009].
- Most of the materials is copy-edited from [Geman, 1990].

Spatial Dependence & Markov Random Fields

- Each pixel value
  - depends only on neighboring pixel values;
  - is independent from far away pixel values.
- Markov random fields provide a flexible mechanism for
  - modeling spatial dependence,
  - image attributes.
- General references for this section are
  i [Geman, 1990].
  ii [Wrinkler, 1995].
  iii [Li, 2009].
- Most of the materials is copy-edited from [Geman, 1990].

Spatial Dependence & Markov Random Fields

- Each pixel value
  - depends only on neighboring pixel values;
  - is independent from far away pixel values.
- Markov random fields provide a flexible mechanism for
  - modeling spatial dependence,
  - image attributes.
- General references for this section are
  i [Geman, 1990].
  ii [Wrinkler, 1995].
  iii [Li, 2009].
- Most of the materials is copy-edited from [Geman, 1990].

Spatial Dependence & Markov Random Fields

- Each pixel value
  - depends only on neighboring pixel values;
  - is independent from far away pixel values.
- Markov random fields provide a flexible mechanism for
  - modeling spatial dependence,
  - image attributes.
- General references for this section are
  i [Geman, 1990].
  ii [Wrinkler, 1995].
  iii [Li, 2009].
- Most of the materials is copy-edited from [Geman, 1990].

Spatial Dependence & Markov Random Fields

- Each pixel value
  - depends only on neighboring pixel values;
  - is independent from far away pixel values.
- Markov random fields provide a flexible mechanism for
  - modeling spatial dependence,
  - image attributes.
- General references for this section are
  i [Geman, 1990].
  ii [Wrinkler, 1995].
  iii [Li, 2009].
- Most of the materials is copy-edited from [Geman, 1990].

Spatial Dependence & Markov Random Fields

- Each pixel value
  - depends only on neighboring pixel values;
  - is independent from far away pixel values.
- Markov random fields provide a flexible mechanism for
  - modeling spatial dependence,
  - image attributes.
- General references for this section are
  i [Geman, 1990].
  ii [Wrinkler, 1995].
  iii [Li, 2009].
- Most of the materials is copy-edited from [Geman, 1990].

Spatial Dependence & Markov Random Fields

- Each pixel value
  - depends only on neighboring pixel values;
  - is independent from far away pixel values.
- Markov random fields provide a flexible mechanism for
  - modeling spatial dependence,
  - image attributes.
- General references for this section are
  i [Geman, 1990].
  ii [Wrinkler, 1995].
  iii [Li, 2009].
- Most of the materials is copy-edited from [Geman, 1990].

State Space

- Given a finite site set S, let

$$x_s, \qquad s \in S \tag{1}$$

be variables indexed by elements of S and belonging to a `state space` Λ_s.
- The state space Λ_s is problem dependent
- State spaces Λ_s may be different from each other.
- **In the following , all the state spaces are assumed to be the same to avoid notational complexity**.

State Space

- Given a finite site set S, let

$$x_s, \qquad s \in S \tag{1}$$

  be variables indexed by elements of S and belonging to a `state space` Λ_s.
- The state space Λ_s is problem dependent
- State spaces Λ_s may be different from each other.
- **In the following , all the state spaces are assumed to be the same to avoid notational complexity**.

State Space

- Given a finite site set S, let

$$x_s, \qquad s \in S \tag{1}$$

  be variables indexed by elements of S and belonging to a `state space` Λ_s.
- The state space Λ_s is problem dependent
- State spaces Λ_s may be different from each other.
- **In the following , all the state spaces are assumed to be the same to avoid notational complexity.**

State Space

- Given a finite site set S, let

$$x_s, \qquad s \in S \tag{1}$$

  be variables indexed by elements of S and belonging to a `state space` Λ_s.
- The state space Λ_s is problem dependent
- State spaces Λ_s may be different from each other.
- **In the following , all the state spaces are assumed to be the same to avoid notational complexity.**

Configuration Space

- Let

$$\Omega = \prod_{s \in S} \Lambda_s = \prod_{s \in S} \Lambda = \Lambda^S. \tag{2}$$

- A map from S to Ω:

$$x : S \longrightarrow \Lambda_s, \tag{3}$$

$$s \longmapsto x(s) = x_s, \tag{4}$$

is called a `configuration` on S with the `configuration space` Ω.

Configuration Space

- Let

$$\Omega = \prod_{s \in S} \Lambda_s = \prod_{s \in S} \Lambda = \Lambda^S. \tag{2}$$

- A map from S to Ω:

$$x : S \longrightarrow \Lambda_s, \tag{3}$$
$$s \longmapsto x(s) = x_s, \tag{4}$$

is called a `configuration` on S with the `configuration space` Ω.

Measure Space

- What is called a `measure`?
- Read Rudin's book, Chapter 1.
- References
  - [Rudin, 1970].

Measure Space

- What is called a `measure`?
- Read Rudin's book, Chapter 1.
- References
  - [Rudin, 1970].

Measure Space

- What is called a `measure`?
- Read Rudin's book, Chapter 1.
- References
  - [Rudin, 1970].

Measure Space

- What is called a `measure`?
- Read Rudin's book, Chapter 1.
- References
  - i [Rudin, 1970].

Measure on State Space

- Assume that there is a positive measure defined on each state space Λ, respectively, i.e.,
  - $(\Lambda, \mathcal{E})$ is a measurable space with positive measure κ on the σ-algebra $\mathcal{E}$.
- The state space Λ is generally a subset of $\mathbf{R}^q$.
- Two typical cases are
  - if Λ is not of zero measure, $\mathcal{E}$ is the Borel algebra and κ some Borel measure;
  - if Λ is a finite or countable subset of $\mathbf{R}^q$, $\mathcal{E}$ is the subset algebra and κ the counting measure.

Measure on State Space

- Assume that there is a positive `measure` defined on each state space Λ, respectively, i.e.,
  - $(\Lambda, \mathcal{E})$ is a measurable space with positive measure κ on the σ-algebra $\mathcal{E}$.
- The state space Λ is generally a subset of $\mathbf{R}^q$.
- Two typical cases are
  - if Λ is not of zero measure, $\mathcal{E}$ is the `Borel algebra` and κ some `Borel measure`;
  - if Λ is a finite or countable subset of $\mathbf{R}^q$, $\mathcal{E}$ is the `subset algebra` and κ the `counting measure`.

Measure on State Space

- Assume that there is a positive `measure` defined on each state space Λ, respectively, i.e.,
  - $(\Lambda, \mathcal{E})$ is a measurable space with positive measure κ on the σ-algebra $\mathcal{E}$.
- The state space Λ is generally a subset of $\mathbf{R}^q$.
- Two typical cases are
  - if Λ is not of zero measure, $\mathcal{E}$ is the `Borel algebra` and κ some `Borel measure`;
  - if Λ is a finite or countable subset of $\mathbf{R}^q$, $\mathcal{E}$ is the `subset algebra` and κ the `counting measure`.

Measure on State Space

- Assume that there is a positive `measure` defined on each state space Λ, respectively, i.e.,
  - $(\Lambda, \mathcal{E})$ is a measurable space with positive measure κ on the σ-algebra $\mathcal{E}$.
- The state space Λ is generally a subset of $\mathbf{R}^q$.
- Two typical cases are
  - if Λ is not of zero measure, $\mathcal{E}$ is the `Borel algebra` and κ some `Borel measure`;
  - if Λ is a finite or countable subset of $\mathbf{R}^q$, $\mathcal{E}$ is the `subset algebra` and κ the `counting measure`.

Measure on State Space

- Assume that there is a positive `measure` defined on each state space Λ, respectively, i.e.,
  - $(\Lambda, \mathcal{E})$ is a measurable space with positive measure κ on the σ-algebra $\mathcal{E}$.
- The state space Λ is generally a subset of $\mathbf{R}^q$.
- Two typical cases are
  - if Λ is not of zero measure, $\mathcal{E}$ is the `Borel algebra` and κ some `Borel measure`;
  - if Λ is a finite or countable subset of $\mathbf{R}^q$, $\mathcal{E}$ is the `subset algebra` and κ the `counting measure`.

Measure on State Space

- Assume that there is a positive `measure` defined on each state space Λ, respectively, i.e.,
  - $(\Lambda, \mathcal{E})$ is a measurable space with positive measure κ on the σ-algebra $\mathcal{E}$.
- The state space Λ is generally a subset of $\mathbf{R}^q$.
- Two typical cases are
  - if Λ is not of zero measure, $\mathcal{E}$ is the `Borel algebra` and κ some `Borel measure`;
  - if Λ is a finite or countable subset of $\mathbf{R}^q$, $\mathcal{E}$ is the `subset algebra` and κ the `counting measure`.

Product σ-algebra on the Configuration Space

- The product σ-algebra

$$\mathcal{T} = \mathcal{E}^S, \tag{5}$$

can be introduced to the configuration space Ω to make it into a measurable space.

- Any element T of $\mathcal{T}$ is called an event.

Product σ-algebra on the Configuration Space

- The `product σ-algebra`

$$\mathcal{T} = \mathcal{E}^S, \tag{5}$$

  can be introduced to the configuration space Ω to make it into a measurable space.

- Any element T of $\mathcal{T}$ is called an event.

Random Fields

- A probability measure on $\mathcal{T}$ defines a `random field`:

Definition

Let S be a finite site set and $(\Lambda, \mathcal{E}, \kappa)$ be a state space. The triple $(\Omega, \mathcal{T}, \Pi)$ is called a random field with the site set S and state space Λ if:

- $(\Omega, \mathcal{T}) = (\Lambda, \mathcal{E})^S$;
- Π *is a probability measure*
  - *a positive measure such that* $\Pi(\Omega) = 1$.

Random Fields

- A probability measure on $\mathcal{T}$ defines a `random field`:

Definition

Let S be a finite site set and $(\Lambda, \mathcal{E}, \kappa)$ be a state space. The triple $(\Omega, \mathcal{T}, \Pi)$ is called a random field with the site set S and state space Λ if:

- $(\Omega, \mathcal{T}) = (\Lambda, \mathcal{E})^S$;
- Π *is a probability measure*
  - *a positive measure such that* $\Pi(\Omega) = 1$.

Random Fields

- A probability measure on $\mathcal{T}$ defines a `random field`:

Definition

Let S be a finite site set and $(\Lambda, \mathcal{E}, \kappa)$ be a state space. The triple $(\Omega, \mathcal{T}, \Pi)$ is called a random field with the site set S and state space Λ if:

- $(\Omega, \mathcal{T}) = (\Lambda, \mathcal{E})^S$;
- *Π is a probability measure*
  - *a positive measure such that $\Pi(\Omega) = 1$.*

Random Fields

- A probability measure on $\mathcal{T}$ defines a `random field`:

Definition

Let S be a finite site set and $(\Lambda, \mathcal{E}, \kappa)$ be a state space. The triple $(\Omega, \mathcal{T}, \Pi)$ is called a random field with the site set S and state space Λ if:

- $(\Omega, \mathcal{T}) = (\Lambda, \mathcal{E})^S$;
- Π *is a probability measure*
  - *a positive measure such that* $\Pi(\Omega) = 1$.

Random Fields

- A probability measure on $\mathcal{T}$ defines a `random field`:

Definition
Let S be a finite site set and $(\Lambda, \mathcal{E}, \kappa)$ be a state space. The triple $(\Omega, \mathcal{T}, \Pi)$ is called a random field with the site set S and state space Λ if:

- $(\Omega, \mathcal{T}) = (\Lambda, \mathcal{E})^S$;
- *Π is a probability measure*
  - *a positive measure such that $\Pi(\Omega) = 1$.*

Discrete and Continuous Random Fields

- If the state space Λ is finite or countable, it is `discrete`.
- If Λ is not of zero measure for the Borel measure on $\mathbf{R}^q$, it is `continuous`.

Discrete and Continuous Random Fields

- If the state space Λ is finite or countable, it is `discrete`.
- If Λ is not of zero measure for the Borel measure on $\mathbf{R}^q$, it is `continuous`.

Coordinate Random Variables

- For any site s, the coordinate random variable X_s with values in Λ is defined as:

$$X_s : (\Omega, \mathcal{T}, \Pi) \longrightarrow (\Lambda, \mathcal{E}) \quad (6)$$

$$x \longmapsto X_s(x) = x_s \quad (7)$$

- To simplify, $X = \{X_s, s \in S\}$.

Coordinate Random Variables

- For any site s, the `coordinate random variable` X_s with values in Λ is defined as:

$$X_s : (\Omega, \mathcal{T}, \Pi) \longrightarrow (\Lambda, \mathcal{E}) \tag{6}$$

$$x \longmapsto X_s(x) = x_s \tag{7}$$

- To simplify, $X = \{X_s, s \in S\}$.

Countable Configuration Assumption

- The sites are sometimes denoted by

$$S = \{s_1, \cdots, s_N\}, \tag{8}$$

where $N = |S|$.

- Configurations $x : S \to \Omega$ are written as

$$x = (x_s), \quad \text{or} \quad x = (x_1, \cdots, x_N) \tag{9}$$

for convenience, with

$$x_i \in \Lambda_{s_i}, \qquad 1 \le i \le N. \tag{10}$$

- Assume each state space Λ_{s_i} is countable in the following.

Countable Configuration Assumption

- The `sites` are sometimes denoted by

$$S = \{s_1, \cdots, s_N\}, \tag{8}$$

where $N = |S|$.

- Configurations $x: S \to \Omega$ are written as

$$x = (x_s), \quad \text{or} \quad x = (x_1, \cdots, x_N) \tag{9}$$

for convenience, with

$$x_i \in \Lambda_{s_i}, \qquad 1 \le i \le N. \tag{10}$$

- Assume each state space Λ_{s_i} is `countable` in the following.

Countable Configuration Assumption

- The `sites` are sometimes denoted by

$$S = \{s_1, \cdots, s_N\}, \tag{8}$$

where $N = |S|$.

- Configurations $x : S \to \Omega$ are written as

$$x = (x_s), \quad \text{or} \quad x = (x_1, \cdots, x_N) \tag{9}$$

for convenience, with

$$x_i \in \Lambda_{s_i}, \qquad 1 \leq i \leq N. \tag{10}$$

- Assume each state space Λ_{s_i} is `countable` in the following.

Conditional Probabilities

- Let $(\Omega, \mathcal{T}, \Pi)$ be a random field with S.
- Assume that Π is a probability measure on Ω with

$$\Pi(x) > 0, \qquad \forall x \in \Omega. \tag{11}$$

- The conditional probabilities

$$\mathbf{Pr}(X_s = x_s, s \in A | X_r = x_r, r \in S \setminus A). \tag{12}$$

where $A \subset S$ are well-defined.
- In the following we simply write it as

$$\Pi(x_s, s \in A | x_r, r \notin A). \tag{13}$$

Conditional Probabilities

- Let $(\Omega, \mathcal{T}, \Pi)$ be a random field with S.
- Assume that Π is a probability measure on Ω with

$$\Pi(x) > 0, \qquad \forall x \in \Omega. \tag{11}$$

- The conditional probabilities

$$\mathbf{Pr}(X_s = x_s, s \in A | X_r = x_r, r \in S \setminus A). \tag{12}$$

where $A \subset S$ are well-defined.

- In the following we simply write it as

$$\Pi(x_s, s \in A | x_r, r \notin A). \tag{13}$$

Conditional Probabilities

- Let $(\Omega, \mathcal{T}, \Pi)$ be a random field with S.
- Assume that Π is a probability measure on Ω with

$$\Pi(x) > 0, \qquad \forall x \in \Omega. \tag{11}$$

- The conditional probabilities

$$\mathbf{Pr}(X_s = x_s, s \in A | X_r = x_r, r \in S \setminus A). \tag{12}$$

where $A \subset S$ are well-defined.

- In the following we simply write it as

$$\Pi(x_s, s \in A | x_r, r \notin A). \tag{13}$$

Conditional Probabilities

- Let $(\Omega, \mathcal{T}, \Pi)$ be a random field with S.
- Assume that Π is a probability measure on Ω with

$$\Pi(x) > 0, \qquad \forall x \in \Omega. \tag{11}$$

- The conditional probabilities

$$\mathbf{Pr}(X_s = x_s, s \in A | X_r = x_r, r \in S \setminus A). \tag{12}$$

where $A \subset S$ are well-defined.
- In the following we simply write it as

$$\Pi(x_s, s \in A | x_r, r \notin A). \tag{13}$$

Local Characteristics

- The `local characteristics` refer to the family of uni-variable, conditional distributions, for $s \in S$ and $x \in \Omega$, and $\lambda \in \Lambda$,

$$\Pi(\lambda | x_{(s)}) \triangleq \Pi_s(x_s | x_{(s)}) \quad (14)$$

$$= \mathbf{Pr}(X_s = x_s | X_r = x_r, r \neq s), \quad (15)$$

where $\lambda = x_s$ and $x_{(s)} = (x_r)_{r \neq s}$.

- **Theorem**

The distribution Π of the random field $(\Omega, \mathcal{T}, \Pi)$ is determined by its local characteristics.

Local Characteristics

- The `local characteristics` refer to the family of uni-variable, conditional distributions, for $s \in S$ and $x \in \Omega$, and $\lambda \in \Lambda$,

$$\Pi(\lambda | x_{(s)}) \triangleq \Pi_s(x_s | x_{(s)}) \quad (14)$$
$$= \mathbf{Pr}(X_s = x_s | X_r = x_r, r \neq s), \quad (15)$$

where $\lambda = x_s$ and $x_{(s)} = (x_r)_{r \neq s}$.

- **Theorem**
The distribution Π of the random field $(\Omega, \mathcal{T}, \Pi)$ is determined by its local characteristics.

Local Characteristics

- The `local characteristics` refer to the family of uni-variable, conditional distributions, for $s \in S$ and $x \in \Omega$, and $\lambda \in \Lambda$,

$$\Pi(\lambda | x_{(s)}) \triangleq \Pi_s(x_s | x_{(s)}) \quad (14)$$

$$= \mathbf{Pr}(X_s = x_s | X_r = x_r, r \neq s), \quad (15)$$

where $\lambda = x_s$ and $x_{(s)} = (x_r)_{r \neq s}$.

- Theorem

The distribution Π of the random field $(\Omega, \mathcal{T}, \Pi)$ is determined by its local characteristics.

Proof I

- We will verify that for any $x = (x_i)$ and $y = (y_i)$,

$$\frac{\Pi(x)}{\Pi(y)} = \prod_{i=1}^{N} \frac{\Pi(x_i | x_1, \cdots, x_{i-1}, y_{i+1}, \cdots, y_N)}{\Pi(y_i | x_1, \cdots, x_{i-1}, y_{i+1}, \cdots, y_N)}. \quad (16)$$

- Assume (16) holds and that two probability measures Π and μ have the same local characteristics.
- It implies that

$$\frac{\Pi(x)}{\Pi(y)} = \frac{\mu(x)}{\mu(y)}. \quad (17)$$

- It follows that

$$\Pi(x)\mu(y) = \mu(x)\Pi(y), \quad (18)$$

- Summing over $y \in \Omega$ leads to the result $\Pi = \mu$.

Proof I

- We will verify that for any $x = (x_i)$ and $y = (y_i)$,

$$\frac{\Pi(x)}{\Pi(y)} = \prod_{i=1}^{N} \frac{\Pi(x_i|x_1, \cdots, x_{i-1}, y_{i+1}, \cdots, y_N)}{\Pi(y_i|x_1, \cdots, x_{i-1}, y_{i+1}, \cdots, y_N)}. \quad (16)$$

- Assume (16) holds and that two probability measures Π and μ have the same local characteristics.
- It implies that

$$\frac{\Pi(x)}{\Pi(y)} = \frac{\mu(x)}{\mu(y)}. \quad (17)$$

- It follows that

$$\Pi(x)\mu(y) = \mu(x)\Pi(y), \quad (18)$$

- Summing over $y \in \Omega$ leads to the result $\Pi = \mu$.

Proof I

- We will verify that for any $x = (x_i)$ and $y = (y_i)$,

$$\frac{\Pi(x)}{\Pi(y)} = \prod_{i=1}^{N} \frac{\Pi(x_i | x_1, \cdots, x_{i-1}, y_{i+1}, \cdots, y_N)}{\Pi(y_i | x_1, \cdots, x_{i-1}, y_{i+1}, \cdots, y_N)}. \qquad (16)$$

- Assume (16) holds and that two probability measures Π and μ have the same local characteristics.
- It implies that

$$\frac{\Pi(x)}{\Pi(y)} = \frac{\mu(x)}{\mu(y)}. \qquad (17)$$

- It follows that

$$\Pi(x)\mu(y) = \mu(x)\Pi(y), \qquad (18)$$

- Summing over $y \in \Omega$ leads to the result $\Pi = \mu$.

Proof I

- We will verify that for any $x = (x_i)$ and $y = (y_i)$,

$$\frac{\Pi(x)}{\Pi(y)} = \prod_{i=1}^{N} \frac{\Pi(x_i|x_1, \cdots, x_{i-1}, y_{i+1}, \cdots, y_N)}{\Pi(y_i|x_1, \cdots, x_{i-1}, y_{i+1}, \cdots, y_N)}. \tag{16}$$

- Assume (16) holds and that two probability measures Π and μ have the same local characteristics.
- It implies that

$$\frac{\Pi(x)}{\Pi(y)} = \frac{\mu(x)}{\mu(y)}. \tag{17}$$

- It follows that

$$\Pi(x)\mu(y) = \mu(x)\Pi(y), \tag{18}$$

- Summing over $y \in \Omega$ leads to the result $\Pi = \mu$.

Proof I

- We will verify that for any $x = (x_i)$ and $y = (y_i)$,

$$\frac{\Pi(x)}{\Pi(y)} = \prod_{i=1}^{N} \frac{\Pi(x_i|x_1, \cdots, x_{i-1}, y_{i+1}, \cdots, y_N)}{\Pi(y_i|x_1, \cdots, x_{i-1}, y_{i+1}, \cdots, y_N)}. \qquad (16)$$

- Assume (16) holds and that two probability measures Π and μ have the same local characteristics.
- It implies that

$$\frac{\Pi(x)}{\Pi(y)} = \frac{\mu(x)}{\mu(y)}. \qquad (17)$$

- It follows that

$$\Pi(x)\mu(y) = \mu(x)\Pi(y), \qquad (18)$$

- Summing over $y \in \Omega$ leads to the result $\Pi = \mu$.

Proof II

- To prove (16), note

$$\Pi(x) = \Pi(x_N|x_1, \cdots, x_{N-1})\Pi(x_1, \cdots, x_{N-1}),$$
$$\Pi(x_1, \cdots, x_{N-1}, y_N) = \Pi(y_N|x_1, \cdots, x_{N-1})\Pi(x_1, \cdots, x_{N-1}).$$

- Therefore

$$\Pi(x) = \Pi(x_1, \cdots, x_{N-1}, x_N) = \frac{\Pi(x_N|x_1, \cdots, x_{N-1})}{\Pi(y_N|x_1, \cdots, x_{N-1})}\Pi(x_1, \cdots, x_{N-1}, y_N).$$

- Similarly

$$\Pi(x_1, \cdots, x_{N-1}, y_N) = \frac{\Pi(x_{N-1}|x_1, \cdots, x_{N-2}, y_N)}{\Pi(y_{N-1}|x_1, \cdots, x_{N-2}, y_N)}\Pi(x_1, \cdots, x_{N-2}, y_{N-1}, y_N).$$

Proof II

- To prove (16), note

$$\Pi(x) = \Pi(x_N|x_1,\cdots,x_{N-1})\Pi(x_1,\cdots,x_{N-1}),$$
$$\Pi(x_1,\cdots,x_{N-1},y_N) = \Pi(y_N|x_1,\cdots,x_{N-1})\Pi(x_1,\cdots,x_{N-1}).$$

- Therefore

$$\Pi(x) = \Pi(x_1,\cdots,x_{N-1},x_N) = \frac{\Pi(x_N|x_1,\cdots,x_{N-1})}{\Pi(y_N|x_1,\cdots,x_{N-1})}\Pi(x_1,\cdots,x_{N-1},y_N).$$

- Similarly

$$\Pi(x_1,\cdots,x_{N-1},y_N) = \frac{\Pi(x_{N-1}|x_1,\cdots,x_{N-2},y_N)}{\Pi(y_{N-1}|x_1,\cdots,x_{N-2},y_N)}\Pi(x_1,\cdots,x_{N-2},y_{N-1},y_N).$$

Proof II

- To prove (16), note

$$\Pi(x) = \Pi(x_N|x_1,\cdots,x_{N-1})\Pi(x_1,\cdots,x_{N-1}),$$
$$\Pi(x_1,\cdots,x_{N-1},y_N) = \Pi(y_N|x_1,\cdots,x_{N-1})\Pi(x_1,\cdots,x_{N-1}).$$

- Therefore

$$\Pi(x) = \Pi(x_1,\cdots,x_{N-1},x_N) = \frac{\Pi(x_N|x_1,\cdots,x_{N-1})}{\Pi(y_N|x_1,\cdots,x_{N-1})}\Pi(x_1,\cdots,x_{N-1},y_N).$$

- Similarly

$$\Pi(x_1,\cdots,x_{N-1},y_N) = \frac{\Pi(x_{N-1}|x_1,\cdots,x_{N-2},y_N)}{\Pi(y_{N-1}|x_1,\cdots,x_{N-2},y_N)}\Pi(x_1,\cdots,x_{N-2},y_{N-1},y_N).$$

Proof III

- $\forall i, 1 < i < N$,

$$\begin{aligned} &\Pi(x_1, \cdots, x_{i-1}, x_i, y_{i+1}, \cdots, y_N) \\ &= \frac{\Pi(x_i | x_1, \cdots, x_{i-1}, y_{i+1}, \cdots, y_N)}{\Pi(y_i | x_1, \cdots, x_{i-1}, y_{i+1}, \cdots, y_N)} \Pi(x_1, \cdots, x_{i-1}, y_i, y_{i+1}, \cdots, y_N). \end{aligned}$$

- For $i = 1$, i.e.,

$$\Pi(x_1, y_2, \cdots, y_N) = \frac{\Pi(x_1 | y_2, \cdots, y_N)}{\Pi(y_1 | y_2, \cdots, y_N)} \Pi(y_1, y_2, \cdots, y_N).$$

- Hence (16) follows by multiplication of the above N equations. $\square$

Proof III

- $\forall i, 1 < i < N,$

$$\Pi(x_1, \cdots, x_{i-1}, x_i, y_{i+1}, \cdots, y_N) = \frac{\Pi(x_i | x_1, \cdots, x_{i-1}, y_{i+1}, \cdots, y_N)}{\Pi(y_i | x_1, \cdots, x_{i-1}, y_{i+1}, \cdots, y_N)} \Pi(x_1, \cdots, x_{i-1}, y_i, y_{i+1}, \cdots, y_N).$$

- For $i = 1$, i.e.,

$$\Pi(x_1, y_2, \cdots, y_N) = \frac{\Pi(x_1 | y_2, \cdots, y_N)}{\Pi(y_1 | y_2, \cdots, y_N)} \Pi(y_1, y_2, \cdots, y_N).$$

- Hence (16) follows by multiplication of the above N equations. □

Proof III

- $\forall i, 1 < i < N,$

$$\Pi(x_1, \cdots, x_{i-1}, x_i, y_{i+1}, \cdots, y_N)$$
$$= \frac{\Pi(x_i | x_1, \cdots, x_{i-1}, y_{i+1}, \cdots, y_N)}{\Pi(y_i | x_1, \cdots, x_{i-1}, y_{i+1}, \cdots, y_N)} \Pi(x_1, \cdots, x_{i-1}, y_i, y_{i+1}, \cdots, y_N).$$

- For $i = 1$, i.e.,

$$\Pi(x_1, y_2, \cdots, y_N) = \frac{\Pi(x_1 | y_2, \cdots, y_N)}{\Pi(y_1 | y_2, \cdots, y_N)} \Pi(y_1, y_2, \cdots, y_N).$$

- Hence (16) follows by multiplication of the above N equations. □

Outline

Neighborhood Systems

- A neighborhood system on S is a collection of subsets $\mathcal{G} = (\mathcal{G}_s)$, $s \in S$, such that

$$\mathcal{G}_s \subset S, \quad \text{if } s \notin \mathcal{G}_s, \tag{19}$$

$$s \in \mathcal{G}_t \iff t \in \mathcal{G}_s. \tag{20}$$

- The pair $(S, \mathcal{G})$ is then a graph:
  - vertexes: sites $s \in S$;
  - edges: pairs $< s, t >$ when $s \in \mathcal{G}_t$.

Neighborhood Systems

- A neighborhood system on S is a collection of subsets $\mathcal{G} = (\mathcal{G}_s)$, $s \in S$, such that

$$\mathcal{G}_s \subset S, \quad \text{if } s \notin \mathcal{G}_s, \tag{19}$$

$$s \in \mathcal{G}_t \Longleftrightarrow t \in \mathcal{G}_s. \tag{20}$$

- The pair $(S, \mathcal{G})$ is then a graph:
  - vertexes: sites $s \in S$;
  - edges: pairs $< s, t >$ when $s \in \mathcal{G}_t$.

Neighborhood Systems

- A neighborhood system on S is a collection of subsets $\mathcal{G} = (\mathcal{G}_s)$, $s \in S$, such that

$$\mathcal{G}_s \subset S, \quad \text{if } s \notin \mathcal{G}_s, \tag{19}$$

$$s \in \mathcal{G}_t \Longleftrightarrow t \in \mathcal{G}_s. \tag{20}$$

- The pair $(S, \mathcal{G})$ is then a graph:
  - vertexes: sites $s \in S$;
  - edges: pairs $< s, t >$ when $s \in \mathcal{G}_t$.

Neighborhood Systems

- A `neighborhood system` on S is a collection of subsets $\mathcal{G} = (\mathcal{G}_s)$, $s \in S$, such that

$$\mathcal{G}_s \subset S, \quad \text{if } s \notin \mathcal{G}_s, \tag{19}$$

$$s \in \mathcal{G}_t \Longleftrightarrow t \in \mathcal{G}_s. \tag{20}$$

- The pair $(S, \mathcal{G})$ is then a `graph`:
  - `vertexes`: sites $s \in S$;
  - `edges`: pairs $< s, t >$ when $s \in \mathcal{G}_t$.

Lattices

- S is usually a subset of d-dimensional `lattice` $\mathbf{Z}^d$.
- Such kinds of site sets are widely used in image processing.
- Bi-dimensional lattices are often used to represent images, each site corresponding to a pixel.
- Tri-dimensional lattices are used to represent 3D data.

Lattices

- S is usually a subset of d-dimensional `lattice` $\mathbf{Z}^d$.
- Such kinds of site sets are widely used in image processing.
- Bi-dimensional lattices are often used to represent images, each site corresponding to a pixel.
- Tri-dimensional lattices are used to represent 3D data.

Lattices

- S is usually a subset of d-dimensional `lattice` $\mathbf{Z}^d$.
- Such kinds of site sets are widely used in image processing.
- Bi-dimensional lattices are often used to represent images, each site corresponding to a pixel.
- Tri-dimensional lattices are used to represent 3D data.

Lattices

- S is usually a subset of d-dimensional `lattice` $\mathbf{Z}^d$.
- Such kinds of site sets are widely used in image processing.
- Bi-dimensional lattices are often used to represent images, each site corresponding to a pixel.
- Tri-dimensional lattices are used to represent 3D data.

Distances on Lattices

- Neighborhood systems can be defined by introducing distances on lattices.
- Some widely used distances:
  - the Euclidean distance;
  - the l^p-distance, $1 \le p \le \infty$:

$$D_p(x, y) = \begin{cases} \left(\sum_d |x_i - y_i|^p\right)^{\frac{1}{p}}, & \text{if } 1 \le p < \infty; \\ \max_{1 \le i \le d} |x_i - y_i|, & \text{if } p = \infty. \end{cases} \quad (21)$$

Distances on Lattices

- Neighborhood systems can be defined by introducing `distances` on lattices.
- Some widely used distances:
  - the Euclidean distance;
  - the l^p-distance, $1 \leq p \leq \infty$:

$$D_p(x, y) = \begin{cases} \left(\sum_d |x_i - y_i|^p\right)^{\frac{1}{p}}, & \text{if } 1 \leq p < \infty; \\ \max_{1 \leq i \leq d} |x_i - y_i|, & \text{if } p = \infty. \end{cases} \tag{21}$$

Distances on Lattices

- Neighborhood systems can be defined by introducing `distances` on lattices.
- Some widely used distances:
  - the Euclidean distance;
  - the l^p-distance, $1 \le p \le \infty$:

$$D_p(x,y) = \begin{cases} \left(\sum_d |x_i - y_i|^p\right)^{\frac{1}{p}}, & \text{if } 1 \le p < \infty; \\ \max_{1 \le i \le d} |x_i - y_i|, & \text{if } p = \infty. \end{cases} \quad (21)$$

Distances on Lattices

- Neighborhood systems can be defined by introducing `distances` on lattices.
- Some widely used distances:
  - the Euclidean distance;
  - the l^p-distance, $1 \leq p \leq \infty$:

$$D_p(x, y) = \begin{cases} \left(\sum_d |x_i - y_i|^p\right)^{\frac{1}{p}}, & \text{if } 1 \leq p < \infty; \\ \max_{1 \leq i \leq d} |x_i - y_i|, & \text{if } p = \infty. \end{cases} \tag{21}$$

Neighborhoods by Distances

- 1-st order neighborhood systems: $\mathcal{G}_s^1$ is the set of the nearest neighbors of s,

$$\mathcal{G}_s^1 \triangleq \underset{t:\, t \neq s}{\arg\min}\, D(s,t). \tag{22}$$

- n-th order neighborhood systems: defined by the recurrence:

$$\mathcal{G}_s^{n+1} \triangleq \mathcal{G}_s^n \cup \underset{t:\, t \notin \mathcal{G}_s^n \cup \{s\}}{\arg\min}\, D(s,t). \tag{23}$$

Neighborhoods by Distances

- 1-st order neighborhood systems: $\mathcal{G}_s^1$ is the set of the nearest neighbors of s,

$$\mathcal{G}_s^1 \triangleq \operatorname*{\mathbf{arg\,min}}_{t:\, t \neq s} D(s,t). \tag{22}$$

- n-th order neighborhood systems: defined by the recurrence:

$$\mathcal{G}_s^{n+1} \triangleq \mathcal{G}_s^n \cup \operatorname*{\mathbf{arg\,min}}_{t:\, t \notin \mathcal{G}_s^n \cup \{s\}} D(s,t). \tag{23}$$

Examples

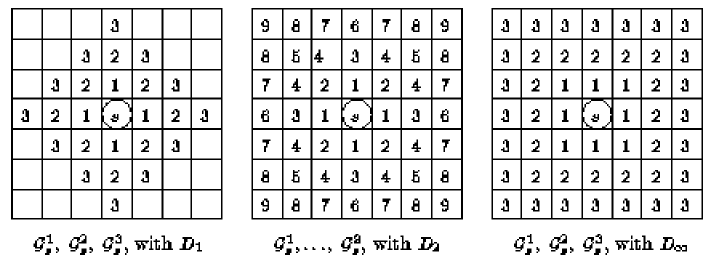

Figure: Neighborhoods of n-th order w.r.t to D_1, D_2, and D_∞ on a 2D regular lattice.

- The most often used are the 1st- and 2nd-order neighborhoods (w.r.t D_2).
- They are called 4-`neighborhood` or 8-`neighborhood` systems in 2D.

References

[Pérez and Heitz, 1994].

Examples

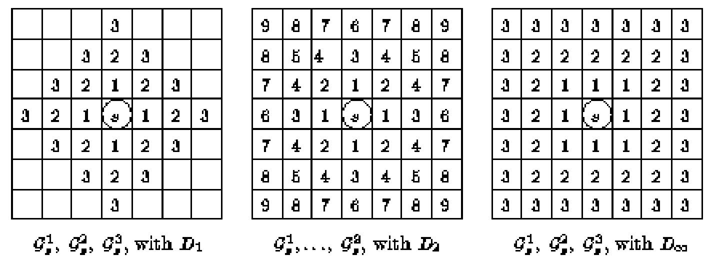

Figure: Neighborhoods of n-th order w.r.t to D_1, D_2, and D_∞ on a 2D regular lattice.

- The most often used are the 1st- and 2nd-order neighborhoods (w.r.t D_2).
- They are called 4`-neighborhood` or 8`-neighborhood` systems in 2D.

References

[Pérez and Heitz, 1994].

Examples

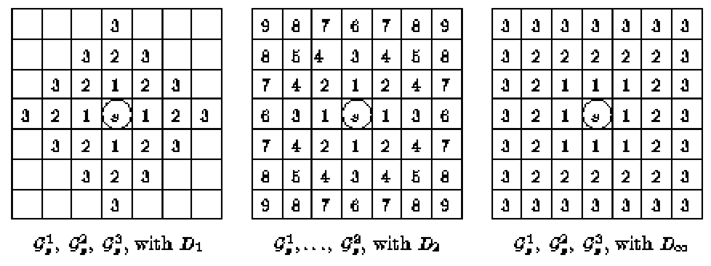

Figure: Neighborhoods of n-th order w.r.t to D_1, D_2, and D_∞ on a 2D regular lattice.

- The most often used are the 1st- and 2nd-order neighborhoods (w.r.t D_2).
- They are called 4`-neighborhood` or 8`-neighborhood` systems in 2D.

References

i [Pérez and Heitz, 1994].

Cliques

- Given a neighborhood system $\mathcal{G} = (\mathcal{G}_s)$, a clique is a set $C \subset S$ if $s, t \in C$ and $s \neq t$, imply $s \in \mathcal{G}_t$.
- Every pair of points are neighbors.
- A single point is a clique.
- The set of all cliques of $\mathcal{G}$ will be denoted by $\mathcal{C}$.

Cliques

- Given a neighborhood system $\mathcal{G} = (\mathcal{G}_s)$, a `clique` is a set $C \subset S$ if $s, t \in C$ and $s \neq t$, imply $s \in \mathcal{G}_t$.
- Every pair of points are neighbors.
- A single point is a clique.
- The set of all cliques of $\mathcal{G}$ will be denoted by $\mathcal{C}$.

Cliques

- Given a neighborhood system $\mathcal{G} = (\mathcal{G}_s)$, a `clique` is a set $C \subset S$ if $s, t \in C$ and $s \neq t$, imply $s \in \mathcal{G}_t$.
- Every pair of points are neighbors.
- A single point is a clique.
- The set of all cliques of $\mathcal{G}$ will be denoted by $\mathcal{C}$.

Cliques

- Given a neighborhood system $\mathcal{G} = (\mathcal{G}_s)$, a `clique` is a set $C \subset S$ if $s, t \in C$ and $s \neq t$, imply $s \in \mathcal{G}_t$.
- Every pair of points are neighbors.
- A single point is a clique.
- The set of all cliques of $\mathcal{G}$ will be denoted by $\mathcal{C}$.

Examples

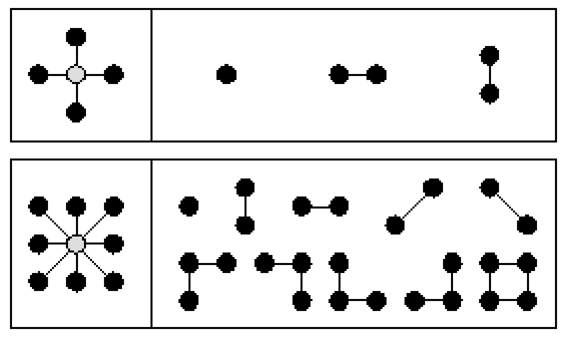

Figure: 1st- and 2nd-order neighborhood systems on a 2D regular lattice (Euclidean distance); associated cliques.

References

- [Pérez and Heitz, 1994].

Outline

Markov Random Fields

- Definition
  A random field $(\Omega, \mathcal{T}, \Pi)$ *with distribution* $\Pi > 0$ *is called a* `Markov random field` *(MRF) with respect to* $\mathcal{G}$ *if*

$$\Pi_s(x_s | x_{(s)}) = \Pi(x_s | x_r, r \in \mathcal{G}_s), \quad \forall s \in S, x \in \Omega. \qquad (24)$$

  - That is,

$$\mathbf{Pr}(X_s = x_s | X_r = x_r, r \neq s) = \mathbf{Pr}(X_s = x_s | X_r = x_r, r \in \mathcal{G}_s). \qquad (25)$$

  - Trivial case:
    - Any probabilistic $\Pi > 0$ defines a MRF w.r.t $\mathcal{G}_s = S - \{s\}$, for $s \in S$.

Markov Random Fields

- Definition
  A random field $(\Omega, \mathcal{T}, \Pi)$ *with distribution* $\Pi > 0$ *is called a* `Markov random field` *(MRF) with respect to* $\mathcal{G}$ *if*

$$\Pi_s(x_s|x_{(s)}) = \Pi(x_s|x_r, r \in \mathcal{G}_s), \quad \forall s \in S, x \in \Omega. \qquad (24)$$

  - That is,

$$\mathbf{Pr}(X_s = x_s|X_r = x_r, r \neq s) = \mathbf{Pr}(X_s = x_s|X_r = x_r, r \in \mathcal{G}_s). \qquad (25)$$

  - Trivial case:
    - Any probabilistic $\Pi > 0$ defines a MRF w.r.t $\mathcal{G}_s = S - \{s\}$, for $s \in S$.

Markov Random Fields

- Definition
  A random field $(\Omega, \mathcal{T}, \Pi)$ *with distribution* $\Pi > 0$ *is called a* `Markov random field` *(MRF) with respect to* $\mathcal{G}$ *if*

$$\Pi_s(x_s | x_{(s)}) = \Pi(x_s | x_r, r \in \mathcal{G}_s), \quad \forall s \in S, x \in \Omega. \tag{24}$$

  - That is,

$$\mathbf{Pr}(X_s = x_s | X_r = x_r, r \neq s) = \mathbf{Pr}(X_s = x_s | X_r = x_r, r \in \mathcal{G}_s). \tag{25}$$

  - Trivial case:
    - Any probabilistic $\Pi > 0$ defines a MRF w.r.t $\mathcal{G}_s = S - \{s\}$, for $s \in S$.

Markov Random Fields

- Definition
  A random field $(\Omega, \mathcal{T}, \Pi)$ *with distribution* $\Pi > 0$ *is called a* `Markov random field` *(MRF) with respect to* $\mathcal{G}$ *if*

$$\Pi_s(x_s | x_{(s)}) = \Pi(x_s | x_r, r \in \mathcal{G}_s), \quad \forall s \in S, x \in \Omega. \qquad (24)$$

  - That is,

$$\mathbf{Pr}(X_s = x_s | X_r = x_r, r \neq s) = \mathbf{Pr}(X_s = x_s | X_r = x_r, r \in \mathcal{G}_s). \qquad (25)$$

  - Trivial case:
    - Any probabilistic $\Pi > 0$ defines a MRF w.r.t $\mathcal{G}_s = S - \{s\}$, for $s \in S$.

Markov Random Fields

- Definition
  A random field $(\Omega, \mathcal{T}, \Pi)$ with distribution $\Pi > 0$ is called a `Markov random field` *(MRF) with respect to $\mathcal{G}$ if*

$$\Pi_s(x_s | x_{(s)}) = \Pi(x_s | x_r, r \in \mathcal{G}_s), \quad \forall s \in S, x \in \Omega. \tag{24}$$

  - That is,

$$\begin{aligned}\mathbf{Pr}(X_s = x_s | X_r = x_r, r \neq s) \\ = \mathbf{Pr}(X_s = x_s | X_r = x_r, r \in \mathcal{G}_s).\end{aligned} \tag{25}$$

  - Trivial case:
    - Any probabilistic $\Pi > 0$ defines a MRF w.r.t $\mathcal{G}_s = S - \{s\}$, for $s \in S$.

Example I

- Let $\{X_n, 0 \leq n \leq N\}$ be a Markov process with state space Λ,
$$P(X_0 = \lambda) = \nu(\lambda) > 0, \tag{26}$$
and transitions
$$P_n(\lambda, \delta) = \mathbf{Pr}(X_{n+1} = \delta | X_n = \lambda) > 0 \tag{27}$$
$\forall \lambda, \delta \in \Lambda$.
- Define
$$\begin{aligned} G_0 &= \{1\}; \\ G_n &= \{n-1, n+1\}, \quad 1 \leq n \leq N-1; \\ G_N &= \{N-1\}. \end{aligned}$$
- Then (X_n) is an MRF with respect to $\mathcal{G} = \{G_j\}_{j=0}^{N}$.

Example I

- Let $\{X_n, 0 \leq n \leq N\}$ be a Markov process with state space Λ,

$$P(X_0 = \lambda) = \nu(\lambda) > 0, \tag{26}$$

and transitions

$$P_n(\lambda, \delta) = \mathbf{Pr}(X_{n+1} = \delta | X_n = \lambda) > 0 \tag{27}$$

$\forall \lambda, \delta \in \Lambda$.

- Define

$$\begin{aligned} G_0 &= \{1\}; \\ G_n &= \{n-1, n+1\}, \quad 1 \leq n \leq N-1; \\ G_N &= \{N-1\}. \end{aligned}$$

- Then (X_n) is an MRF with respect to $\mathcal{G} = \{G_j\}_{j=0}^{N}$.

Example I

- Let $\{X_n, 0 \leq n \leq N\}$ be a Markov process with state space Λ,

$$P(X_0 = \lambda) = \nu(\lambda) > 0, \tag{26}$$

and transitions

$$P_n(\lambda, \delta) = \mathbf{Pr}(X_{n+1} = \delta | X_n = \lambda) > 0 \tag{27}$$

$\forall \lambda, \delta \in \Lambda$.

- Define

$$G_0 = \{1\};$$
$$G_n = \{n-1, n+1\}, \quad 1 \leq n \leq N-1;$$
$$G_N = \{N-1\}.$$

- Then (X_n) is an MRF with respect to $\mathcal{G} = \{G_j\}_{j=0}^{N}$.

Example II

- The local characteristics are

$$\Pi_0(x_0|x_{(0)}) = \frac{\nu(x_0)P_0(x_0, x_1)}{\sum_{\lambda\in\Lambda}\nu(\lambda)P_0(\lambda, x_1)} \tag{28}$$

$$\Pi_n(x_n|x_{(n)}) = \frac{P_{n-1}(x_{n-1}, x_n)P_n(x_n, x_{n+1})}{\sum_{\lambda\in\Lambda}P_{n-1}(x_{n-1}, \lambda)P_n(\lambda, x_{n+1})}, \quad 1 \le n \le N-1 \tag{29}$$

$$\Pi_N(x_N|x_{(N)}) = P_{N-1}(x_{N-1}, x_N). \tag{30}$$

- By Theorem 1.2, the local characteristics determine a unique Markov random field Π on $(\{0, \cdots, N\}, \mathcal{G})$.

Example II

- The local characteristics are

$$\Pi_0(x_0|x_{(0)}) = \frac{\nu(x_0)P_0(x_0, x_1)}{\sum_{\lambda\in\Lambda}\nu(\lambda)P_0(\lambda, x_1)} \quad (28)$$

$$\Pi_n(x_n|x_{(n)}) = \frac{P_{n-1}(x_{n-1}, x_n)P_n(x_n, x_{n+1})}{\sum_{\lambda\in\Lambda}P_{n-1}(x_{n-1}, \lambda)P_n(\lambda, x_{n+1})}, \quad 1 \le n \le N-1 \quad (29)$$

$$\Pi_N(x_N|x_{(N)}) = P_{N-1}(x_{N-1}, x_N). \quad (30)$$

- By Theorem 1.2, the local characteristics determine a unique Markov random field Π on $(\{0, \cdots, N\}, \mathcal{G})$.

Proof I

- By definition, (X_n) is Markov if $\forall m \geq 0$

$$\begin{aligned}\mathbf{Pr}(X_{m+1} = x_{m+1} | X_j = x_j, 0 \leq j \leq m) \\ = \mathbf{Pr}(X_{m+1} = x_{m+1} | X_m = x_m). \quad (31)\end{aligned}$$

- Therefore

$$\begin{aligned}&\mathbf{Pr}(X_0 = x_0, X_1 = x_1, X_2 = x_2, X_3 = x_3) \\ &= \mathbf{Pr}(x_0, x_1, x_2, x_3) = \mathbf{Pr}(x_3 | x_0, x_1, x_2)\mathbf{Pr}(x_0, x_1, x_2) \\ &= \mathbf{Pr}(x_3 | x_2)\mathbf{Pr}(x_2 | x_1)\mathbf{Pr}(x_1 | x_0)\mathbf{Pr}(x_0).\end{aligned}$$

- Generally, we have,

$$\begin{aligned}&\mathbf{Pr}(X_0 = x_0, \cdots, X_m = x_m) = \mathbf{Pr}(x_0, \cdots, x_m) \\ &= \mathbf{Pr}(x_0) \prod_{i=0}^{m-1} \mathbf{Pr}(x_{i+1} | x_i) \quad (32) \\ &= \nu(x_0) \prod_{i=0}^{m-1} P_i(x_i, x_{i+1}).\end{aligned}$$

Proof I

- By definition, (X_n) is Markov if $\forall m \geq 0$

$$\begin{aligned}\mathbf{Pr}(X_{m+1} = x_{m+1}|X_j = x_j, 0 \leq j \leq m) \\ = \mathbf{Pr}(X_{m+1} = x_{m+1}|X_m = x_m). \quad (31)\end{aligned}$$

- Therefore

$$\begin{aligned}&\mathbf{Pr}(X_0 = x_0, X_1 = x_1, X_2 = x_2, X_3 = x_3) \\ &= \mathbf{Pr}(x_0, x_1, x_2, x_3) = \mathbf{Pr}(x_3|x_0, x_1, x_2)\mathbf{Pr}(x_0, x_1, x_2) \\ &= \mathbf{Pr}(x_3|x_2)\mathbf{Pr}(x_2|x_1)\mathbf{Pr}(x_1|x_0)\mathbf{Pr}(x_0).\end{aligned}$$

- Generally, we have,

$$\begin{aligned}\mathbf{Pr}(X_0 = x_0, \cdots, X_m = x_m) = \mathbf{Pr}(x_0, \cdots, x_m) \\ = \mathbf{Pr}(x_0) \prod_{i=0}^{m-1} \mathbf{Pr}(x_{i+1}|x_i) \quad (32) \\ = \nu(x_0) \prod_{i=0}^{m-1} P_i(x_i, x_{i+1}).\end{aligned}$$

Proof I

- By definition, (X_n) is Markov if $\forall m \geq 0$

$$\begin{aligned}\mathbf{Pr}(X_{m+1} = x_{m+1}|X_j = x_j, 0 \leq j \leq m) \\ = \mathbf{Pr}(X_{m+1} = x_{m+1}|X_m = x_m). \quad (31)\end{aligned}$$

- Therefore

$$\begin{aligned}&\mathbf{Pr}(X_0 = x_0, X_1 = x_1, X_2 = x_2, X_3 = x_3) \\ &= \mathbf{Pr}(x_0, x_1, x_2, x_3) = \mathbf{Pr}(x_3|x_0, x_1, x_2)\mathbf{Pr}(x_0, x_1, x_2) \\ &= \mathbf{Pr}(x_3|x_2)\mathbf{Pr}(x_2|x_1)\mathbf{Pr}(x_1|x_0)\mathbf{Pr}(x_0).\end{aligned}$$

- Generally, we have,

$$\begin{aligned}&\mathbf{Pr}(X_0 = x_0, \cdots, X_m = x_m) = \mathbf{Pr}(x_0, \cdots, x_m) \\ &= \mathbf{Pr}(x_0) \prod_{i=0}^{m-1} \mathbf{Pr}(x_{i+1}|x_i) \qquad (32) \\ &= \nu(x_0) \prod_{i=0}^{m-1} P_i(x_i, x_{i+1}).\end{aligned}$$

Proof — case a: $1 \le n \le N-1$ (1)

- By (32)

$$\begin{aligned}\mathbf{Pr}(x_n|x_{(n)}) &= \mathbf{Pr}(x_n|x_0,\cdots,x_{n-1},x_{n+1},\cdots,x_N)\\ &= \frac{\mathbf{Pr}(x_0,\cdots,x_{n-1},x_n,x_{n+1},\cdots,x_N)}{\mathbf{Pr}(x_0,\cdots,x_{n-1},x_{n+1},\cdots,x_N)}\\ &= \frac{\mathbf{Pr}(x_0,\cdots,x_{n-1},x_n,x_{n+1},\cdots,x_N)}{\sum_{\lambda\in\Lambda}\mathbf{Pr}(x_0,\cdots,x_{n-1},\lambda,x_{n+1},\cdots,x_N)}\\ &= \frac{P_{n-1}(x_{n-1},x_n)P_n(x_n,x_{n+1})}{\sum_{\lambda\in\Lambda}P_{n-1}(x_{n-1},\lambda)P_n(\lambda,x_{n+1})};\end{aligned}$$

and

$$\begin{aligned}\mathbf{Pr}(x_n|x_r,r\in G_r) &= \mathbf{Pr}(x_n|x_{n-1},x_{n+1})\\ &= \frac{\mathbf{Pr}(x_{n-1},x_n,x_{n+1})}{\mathbf{Pr}(x_{n-1},x_{n+1})}\\ &= \frac{\mathbf{Pr}(x_{n-1},x_n,x_{n+1})}{\sum_{\lambda\in\Lambda}\mathbf{Pr}(x_{n-1},\lambda,x_{n+1})}.\end{aligned}$$

Proof — case a: $1 \leq n \leq N - 1$ (2)

- Because

$$\mathbf{Pr}(x_{n-1}, x_n, x_{n+1}) = \sum_{\lambda_0,\cdots,\lambda_{n-2}\in\Lambda} \mathbf{Pr}(\lambda_0, \cdots, \lambda_{n-2}, x_{n-1}, x_n, x_{n+1})$$

$$= \sum_{\lambda_0,\cdots,\lambda_{n-2}\in\Lambda} \mathbf{Pr}(x_{n+1}|x_n)\mathbf{Pr}(x_n|x_{n-1})\mathbf{Pr}(\lambda_0, \cdots, \lambda_{n-2}, x_{n-1})$$

$$= \mathbf{Pr}(x_{n+1}|x_n)\mathbf{Pr}(x_n|x_{n-1})\mathbf{Pr}(x_{n-1}),$$

- it follows that

$$\mathbf{Pr}(x_n|x_r, r \in G_r) = \mathbf{Pr}(x_n|x_{(n)}).$$

Proof — case a: $1 \leq n \leq N-1$ (2)

- Because

$$\mathbf{Pr}(x_{n-1}, x_n, x_{n+1}) = \sum_{\lambda_0,\cdots,\lambda_{n-2}\in\Lambda} \mathbf{Pr}(\lambda_0,\cdots,\lambda_{n-2}, x_{n-1}, x_n, x_{n+1})$$

$$= \sum_{\lambda_0,\cdots,\lambda_{n-2}\in\Lambda} \mathbf{Pr}(x_{n+1}|x_n)\mathbf{Pr}(x_n|x_{n-1})\mathbf{Pr}(\lambda_0,\cdots,\lambda_{n-2}, x_{n-1})$$

$$= \mathbf{Pr}(x_{n+1}|x_n)\mathbf{Pr}(x_n|x_{n-1})\mathbf{Pr}(x_{n-1}),$$

- it follows that

$$\mathbf{Pr}(x_n|x_r, r \in G_r) = \mathbf{Pr}(x_n|x_{(n)}).$$

Proof — case b: $n = 0$

- By (32)

$$\mathbf{Pr}(x_0|x_{(0)}) = \mathbf{Pr}(x_0|x_1, \cdots, x_N) = \frac{\mathbf{Pr}(x_0, x_1, \cdots, x_N)}{\mathbf{Pr}(x_1, \cdots, x_N)}$$
$$= \frac{\mathbf{Pr}(x_0, x_1, \cdots, x_N)}{\sum_{\lambda \in \Lambda} \mathbf{Pr}(\lambda, x_1, \cdots, x_N)}$$
$$= \frac{\nu(x_0) P_0(x_0, x_1)}{\sum_{\lambda \in \Lambda} \nu(\lambda) P_0(\lambda, x_1)};$$

- $$\mathbf{Pr}(x_0|x_r, r \in G_0) = \mathbf{Pr}(x_0|x_1) = \frac{\mathbf{Pr}(x_0, x_1)}{\mathbf{Pr}(x_1)} = \frac{\mathbf{Pr}(x_1, x_0)}{\sum_{\lambda \in \Lambda} \mathbf{Pr}(x_1, \lambda)}$$
$$= \frac{\mathbf{Pr}(x_1|x_0)\mathbf{Pr}(x_0)}{\sum_{\lambda \in \Lambda} \mathbf{Pr}(x_1|\lambda)\mathbf{Pr}(\lambda)}$$
$$= \frac{\nu(x_0) P_0(x_0, x_1)}{\sum_{\lambda \in \Lambda} \nu(\lambda) P_0(\lambda, x_1)}.$$

Proof — case b: $n = 0$

- By (32)

$$\mathbf{Pr}(x_0|x_{(0)}) = \mathbf{Pr}(x_0|x_1,\cdots,x_N) = \frac{\mathbf{Pr}(x_0,x_1,\cdots,x_N)}{\mathbf{Pr}(x_1,\cdots,x_N)}$$
$$= \frac{\mathbf{Pr}(x_0,x_1,\cdots,x_N)}{\sum_{\lambda\in\Lambda}\mathbf{Pr}(\lambda,x_1,\cdots,x_N)}$$
$$= \frac{\nu(x_0)P_0(x_0,x_1)}{\sum_{\lambda\in\Lambda}\nu(\lambda)P_0(\lambda,x_1)};$$

-

$$\mathbf{Pr}(x_0|x_r,r\in G_0) = \mathbf{Pr}(x_0|x_1) = \frac{\mathbf{Pr}(x_0,x_1)}{\mathbf{Pr}(x_1)} = \frac{\mathbf{Pr}(x_1,x_0)}{\sum_{\lambda\in\Lambda}\mathbf{Pr}(x_1,\lambda)}$$
$$= \frac{\mathbf{Pr}(x_1|x_0)\mathbf{Pr}(x_0)}{\sum_{\lambda\in\Lambda}\mathbf{Pr}(x_1|\lambda)\mathbf{Pr}(\lambda)}$$
$$= \frac{\nu(x_0)P_0(x_0,x_1)}{\sum_{\lambda\in\Lambda}\nu(\lambda)P_0(\lambda,x_1)}.$$

Proof — case c: $n = N$

- By the Markov property of X,

$$\mathbf{Pr}(x_N | x_{(N)}) = \mathbf{Pr}(x_N | x_0, \cdots, x_{N-1}) = \mathbf{Pr}(x_N | x_{N-1})$$

and

$$\mathbf{Pr}(x_N | x_r, r \in G_N) = \mathbf{Pr}(x_N | x_{N-1}).$$

Outline

Potentials

- Gibbsian random fields are representations for positive measures motivated by equilibrium studies in statistical physics.

- A potential is a collection of functions defined on Ω indexed by the subsets of S,

$$V = \{V_A : A \subset S, V_A : \Omega \to \mathbf{R}\}$$

such that

$$V_\emptyset = 0; \qquad (33)$$

$$V_A(x) = V_A(x'), \text{ if } x_s = x'_s \text{ for all } s \in A, \qquad (34)$$

i.e., $V_A(x)$ depends only on those coordinates x_s of x for which $s \in A$.

Potentials

- Gibbsian random fields are representations for positive measures motivated by equilibrium studies in statistical physics.
- A potential is a collection of functions defined on Ω indexed by the subsets of S,

$$V = \{V_A : A \subset S, V_A : \Omega \to \mathbf{R}\}$$

such that

$$V_\emptyset = 0; \quad (33)$$

$$V_A(x) = V_A(x'), \text{ if } x_s = x'_s \text{ for all } s \in A, \quad (34)$$

i.e., $V_A(x)$ depends only on those coordinates x_s of x for which $s \in A$.

Hamiltonian

- V is `normalized` if

$$V_A(x) = 0, \tag{35}$$

whenever $x_t = 0$ for some $t \in A$.

- It is assumed that $0 \in \Lambda_s, \forall s$, although any other distinguished point would do equally well.
- This condition is only imposed to insure unique representations; it has no practical importance.

- The `energy` or `Hamiltonian` associated with V is

$$H(x) = H_V(x) = \sum_{A \subset S} V_A(x). \tag{36}$$

Hamiltonian

- V is `normalized` if

$$V_A(x) = 0, \tag{35}$$

  whenever $x_t = 0$ for some $t \in A$.
- It is assumed that $0 \in \Lambda_s, \forall s$, although any other distinguished point would do equally well.
- This condition is only imposed to insure unique representations; it has no practical importance.
- The `energy` or `Hamiltonian` associated with V is

$$H(x) = H_V(x) = \sum_{A \subset S} V_A(x). \tag{36}$$

Hamiltonian

- V is `normalized` if

$$V_A(x) = 0, \quad (35)$$

whenever $x_t = 0$ for some $t \in A$.

- It is assumed that $0 \in \Lambda_s$, $\forall s$, although any other distinguished point would do equally well.
- This condition is only imposed to insure unique representations; it has no practical importance.
- The `energy` or `Hamiltonian` associated with V is

$$H(x) = H_V(x) = \sum_{A \subset S} V_A(x). \quad (36)$$

Hamiltonian

- V is normalized if

$$V_A(x) = 0, \tag{35}$$

whenever $x_t = 0$ for some $t \in A$.

- It is assumed that $0 \in \Lambda_s, \forall s$, although any other distinguished point would do equally well.
- This condition is only imposed to insure unique representations; it has no practical importance.

- The energy or Hamiltonian associated with V is

$$H(x) = H_V(x) = \sum_{A \subset S} V_A(x). \tag{36}$$

Gibbsian Random Fields

- A Gibbsian random field w.r.t $G = [S, \mathcal{G}]$ is a measure of the form

$$\Pi(x) = Z^{-1}\mathbf{e}^{-H(x)}, \quad Z = \sum_x \mathbf{e}^{-H(x)} \tag{37}$$

such that $Z < +\infty$ if $|\Omega| = \infty$ and

- V is a Gibbsian potential, i.e.,

$$V_A = 0, \quad \forall A \notin \mathcal{C}; \tag{38}$$

- the Hamiltonian

$$H(x) = \sum_{C \in \mathcal{C}} V_C(x); \tag{39}$$

where $\mathcal{C}$ is the set of all cliques of $\mathcal{G}$.

Gibbsian Random Fields

- A Gibbsian random field w.r.t $G = [S, \mathcal{G}]$ is a measure of the form

$$\Pi(x) = Z^{-1}\mathbf{e}^{-H(x)}, \quad Z = \sum_x \mathbf{e}^{-H(x)} \tag{37}$$

such that $Z < +\infty$ if $|\Omega| = \infty$ and

  - V is a Gibbsian potential, i.e.,

$$V_A = 0, \quad \forall A \notin \mathcal{C}; \tag{38}$$

  - the Hamiltonian

$$H(x) = \sum_{C \in \mathcal{C}} V_C(x); \tag{39}$$

where $\mathcal{C}$ is the set of all cliques of $\mathcal{G}$.

Gibbsian Random Fields

- A `Gibbsian random field` w.r.t $G = [S, \mathcal{G}]$ is a measure of the form

$$\Pi(x) = Z^{-1}\mathbf{e}^{-H(x)}, \quad Z = \sum_x \mathbf{e}^{-H(x)} \tag{37}$$

such that $Z < +\infty$ if $|\Omega| = \infty$ and

- V is a `Gibbsian potential`, i.e.,

$$V_A = 0, \quad \forall A \notin \mathcal{C}; \tag{38}$$

- the Hamiltonian

$$H(x) = \sum_{C \in \mathcal{C}} V_C(x); \tag{39}$$

where $\mathcal{C}$ is the set of all cliques of $\mathcal{G}$.

Exponential Family

- With few exceptions, the `partition function` Z is intractable both analytically and numerically.
- Typically, there are parameters $\theta = (\theta_1, \cdots, \theta_J)$ in V, so that

$$Z = Z(\theta) = \sum_{x \in \Omega} \mathbf{e}^{-H(x;\theta)}.$$

- The special case

$$H(x;\theta) = \sum_{j=1}^{J} \theta_j H_j(x), \tag{40}$$

is an example of an `exponential family`.

Exponential Family

- With few exceptions, the `partition function` Z is intractable both analytically and numerically.
- Typically, there are parameters $\theta = (\theta_1, \cdots, \theta_J)$ in V, so that

$$Z = Z(\theta) = \sum_{x \in \Omega} \mathbf{e}^{-H(x;\theta)}.$$

- The special case

$$H(x;\theta) = \sum_{j=1}^{J} \theta_j H_j(x), \tag{40}$$

is an example of an `exponential family`.

Exponential Family

- With few exceptions, the `partition function` Z is intractable both analytically and numerically.
- Typically, there are parameters $\theta = (\theta_1, \cdots, \theta_J)$ in V, so that

$$Z = Z(\theta) = \sum_{x \in \Omega} \mathbf{e}^{-H(x;\theta)}.$$

- The special case

$$H(x;\theta) = \sum_{j=1}^{J} \theta_j H_j(x), \tag{40}$$

is an example of an `exponential family`.

Maximum Entropy Principle and Gibbsian Random Fields

- Assume that the distribution Π of the random variables $\{X_s, s \in S\}$ satisfies the following expectation conditions

$$E[V_C(X,\theta)] = \sum_{x\in\Omega} V_C(x,\theta)\Pi(x) = \mu_C(\theta),\ \forall C \in \mathcal{C}, \tag{41}$$

  where θ is a parameter.
- The maximum entropy principle concludes that

$$\Pi(X_s = x_s, s \in S) = \Pi(x) = \frac{1}{Z}\Pi_0(x)\mathbf{e}^{\sum_{C\in\mathcal{C}} \lambda_k(\theta) V_C(x,\theta)} \tag{42}$$

  where $\Pi_0(x)$ is some *a priori* distribution.
- Hence, Gibbsian random fields can be induced by the maximum entropy principle.

Maximum Entropy Principle and Gibbsian Random Fields

- Assume that the distribution Π of the random variables $\{X_s, s \in S\}$ satisfies the following expectation conditions

$$E[V_C(X,\theta)] = \sum_{x \in \Omega} V_C(x,\theta)\Pi(x) = \mu_C(\theta), \ \forall C \in \mathcal{C}, \tag{41}$$

  where θ is a parameter.
- The maximum entropy principle concludes that

$$\Pi(X_s = x_s, s \in S) = \Pi(x) = \frac{1}{Z}\Pi_0(x)\mathbf{e}^{\sum_{C \in \mathcal{C}} \lambda_k(\theta) V_C(x,\theta)} \tag{42}$$

  where $\Pi_0(x)$ is some *a priori* distribution.
- Hence, Gibbsian random fields can be induced by the maximum entropy principle.

Maximum Entropy Principle and Gibbsian Random Fields

- Assume that the distribution Π of the random variables $\{X_s, s \in S\}$ satisfies the following expectation conditions

$$E[V_C(X,\theta)] = \sum_{x \in \Omega} V_C(x,\theta)\Pi(x) = \mu_C(\theta), \ \forall C \in \mathcal{C}, \tag{41}$$

  where θ is a parameter.
- The maximum entropy principle concludes that

$$\Pi(X_s = x_s, s \in S) = \Pi(x) = \frac{1}{Z}\Pi_0(x)\mathbf{e}^{\sum_{C \in \mathcal{C}} \lambda_k(\theta) V_C(x,\theta)} \tag{42}$$

  where $\Pi_0(x)$ is some *a priori* distribution.
- Hence, Gibbsian random fields can be induced by the maximum entropy principle.

Möbius Inversion Formula

- Let Φ and Ψ be set functions on the power set $\mathcal{P}(S)$, $|S| < \infty$. Then

$$\Phi(A) = \sum_{B \subset A} (-1)^{|A-B|} \Psi(B), \quad \forall A \subset S, \tag{43}$$

if and only if

$$\Psi(A) = \sum_{B \subset A} \Phi(B), \quad \forall A \subset S. \tag{44}$$

- This is used in proving the following representation theorem.

Möbius Inversion Formula

- Let Φ and Ψ be set functions on the power set $\mathcal{P}(S)$, $|S| < \infty$. Then

$$\Phi(A) = \sum_{B \subset A} (-1)^{|A-B|} \Psi(B), \quad \forall A \subset S, \tag{43}$$

if and only if

$$\Psi(A) = \sum_{B \subset A} \Phi(B), \quad \forall A \subset S. \tag{44}$$

- This is used in proving the following representation theorem.

Notation

- For $x \in \Omega$, $A \subset S$, set

$$x^A = (x_s^A), \quad x_s^A = \begin{cases} x_s, & s \in A \\ 0, & s \notin A \end{cases} \tag{45}$$

Representation Theorem

Theorem

For every random field $\Pi > 0$, let

$$V_A(x) = -\sum_{B \subset A} (-1)^{|A-B|} \log \Pi(x^B), \quad (46)$$

and $V_\phi = 0$. Then

$$\Pi(x) = Z^{-1} \mathbf{e}^{-H(x)} \quad (47)$$

where $H(x) = \sum_{B \subset S} V_B(x)$ and $Z = \Pi(0)^{-1}$. Moreover, for any $s \in A$,

$$V_A(x) = -\sum_{B \subset A} (-1)^{|A-B|} \log \Pi(x_s^B | x_{(s)}^B). \quad (48)$$

The representation of V_A is unique among normalized potentials.

Lemma

Lemma

For every finite set A

$$\sum_{B\subset A}(-1)^{|A-B|}=\sum_{B\subset A}(-1)^{|B|}=\begin{cases}1, & \text{if } A=\emptyset;\\ 0, & \text{if } A\neq\emptyset\end{cases} \tag{49}$$

- If $A=\emptyset$, the result is obvious.
- If $A\neq\emptyset$,

$$\sum_{B\subset A}(-1)^{|B|}=\sum_{k=0}^{|A|}|\{B\subset A:|B|=k\}|(-1)^k \tag{50}$$

$$=\sum_{k=0}^{|A|}C_{|A|}^{k}(-1)^k=(1-1)^{|A|}=0. \tag{51}$$

Lemma

Lemma

For every finite set A

$$\sum_{B\subset A}(-1)^{|A-B|} = \sum_{B\subset A}(-1)^{|B|} = \begin{cases} 1, & \textit{if } A=\emptyset; \\ 0, & \textit{if } A\neq\emptyset \end{cases} \tag{49}$$

- If $A=\emptyset$, the result is obvious.
- If $A\neq\emptyset$,

$$\sum_{B\subset A}(-1)^{|B|} = \sum_{k=0}^{|A|}|\{B\subset A : |B|=k\}|(-1)^k \tag{50}$$

$$= \sum_{k=0}^{|A|} C_{|A|}^k(-1)^k = (1-1)^{|A|} = 0. \tag{51}$$

Lemma

Lemma

For every finite set A

$$\sum_{B\subset A}(-1)^{|A-B|}=\sum_{B\subset A}(-1)^{|B|}=\begin{cases}1, & \text{if } A=\emptyset;\\ 0, & \text{if } A\neq\emptyset\end{cases} \tag{49}$$

- If $A=\emptyset$, the result is obvious.
- If $A\neq\emptyset$,

$$\sum_{B\subset A}(-1)^{|B|}=\sum_{k=0}^{|A|}|\{B\subset A:|B|=k\}|(-1)^k \tag{50}$$

$$=\sum_{k=0}^{|A|}C_{|A|}^{k}(-1)^k=(1-1)^{|A|}=0. \tag{51}$$

Proof — step 1

- *the representation of Π in (47) is valid.*
- Define

$$\Psi(A) = -\log\left[\frac{\Pi(x^A)}{\Pi(0)}\right]$$
$$\Phi(A) = V_A(x)$$

where x is fixed and $0 = (0, \cdots, 0)$.
- Assuming (46), by the lemma and using the Möbius inversion formula for Ψ,

$$-\log\left[\frac{\Pi(x)}{\Pi(0)}\right] = -\log\left[\frac{\Pi(x^S)}{\Pi(0)}\right] = \Psi(S) = \sum_{B\subset S} V_B(x).$$

Thus, $\Pi(x) = \Pi(0)\mathbf{e}^{-H(x)}$.

Proof — step 1

- *the representation of Π in (47) is valid*.
- Define

$$\Psi(A) = -\log\left[\frac{\Pi(x^A)}{\Pi(0)}\right]$$

$$\Phi(A) = V_A(x)$$

where x is fixed and $0 = (0, \cdots, 0)$.

- Assuming (46), by the lemma and using the Möbius inversion formula for Ψ,

$$-\log\left[\frac{\Pi(x)}{\Pi(0)}\right] = -\log\left[\frac{\Pi(x^S)}{\Pi(0)}\right] = \Psi(S) = \sum_{B\subset S} V_B(x).$$

Thus, $\Pi(x) = \Pi(0)\mathbf{e}^{-H(x)}$.

Proof — step 1

- *the representation of* Π *in (47) is valid*.
- Define

$$\Psi(A) = -\log\left[\frac{\Pi(x^A)}{\Pi(0)}\right]$$

$$\Phi(A) = V_A(x)$$

where x is fixed and $0 = (0, \cdots, 0)$.

- Assuming (46), by the lemma and using the Möbius inversion formula for Ψ,

$$-\log\left[\frac{\Pi(x)}{\Pi(0)}\right] = -\log\left[\frac{\Pi(x^S)}{\Pi(0)}\right] = \Psi(S) = \sum_{B\subset S} V_B(x).$$

Thus, $\Pi(x) = \Pi(0)\mathbf{e}^{-H(x)}$.

Proof — step 2

- *V is normalized*
- For any $s \in A$,

$$
\begin{aligned}
& - V_A(x) \\
&= \sum_{B \subset A, s \notin B} (-1)^{|A-B|} \log \Pi(x^B) + \sum_{B \subset A, s \in B} (-1)^{|A-B|} \log \Pi(x^B) \\
&= \sum_{B \subset A - \{s\}} (-1)^{|A-B|} \log \Pi(x^B) \\
&\qquad + \sum_{B' \subset A - \{s\}} (-1)^{|A-B'-\{s\}|} \log \Pi(x^{B' \cup \{s\}}) \\
&= \sum_{B \subset A - \{s\}} (-1)^{|A-B|} \left(\log \Pi(x^B) - \log \Pi(x^{B \cup \{s\}}) \right).
\end{aligned}
$$

- If $x_s = 0$, then $x^B = x^{B \cup \{s\}}$.
- Hence $V_A(x) = 0$.

Proof — step 2

- *V is normalized*
- For any $s \in A$,

$$
\begin{aligned}
&- V_A(x) \\
&= \sum_{B \subset A, s \notin B} (-1)^{|A-B|} \log \Pi(x^B) + \sum_{B \subset A, s \in B} (-1)^{|A-B|} \log \Pi(x^B) \\
&= \sum_{B \subset A-\{s\}} (-1)^{|A-B|} \log \Pi(x^B) \\
&\qquad + \sum_{B' \subset A-\{s\}} (-1)^{|A-B'-\{s\}|} \log \Pi(x^{B' \cup \{s\}}) \\
&= \sum_{B \subset A-\{s\}} (-1)^{|A-B|} \left(\log \Pi(x^B) - \log \Pi(x^{B \cup \{s\}}) \right).
\end{aligned}
$$

- If $x_s = 0$, then $x^B = x^{B \cup \{s\}}$.
- Hence $V_A(x) = 0$.

Proof — step 2

- *V is normalized*
- For any $s \in A$,

$$
\begin{aligned}
& - V_A(x) \\
&= \sum_{B \subset A, s \notin B} (-1)^{|A-B|} \log \Pi(x^B) + \sum_{B \subset A, s \in B} (-1)^{|A-B|} \log \Pi(x^B) \\
&= \sum_{B \subset A - \{s\}} (-1)^{|A-B|} \log \Pi(x^B) \\
&\qquad + \sum_{B' \subset A - \{s\}} (-1)^{|A-B'-\{s\}|} \log \Pi(x^{B' \cup \{s\}}) \\
&= \sum_{B \subset A - \{s\}} (-1)^{|A-B|} \left(\log \Pi(x^B) - \log \Pi(x^{B \cup \{s\}}) \right).
\end{aligned}
$$

- If $x_s = 0$, then $x^B = x^{B \cup \{s\}}$.
- Hence $V_A(x) = 0$.

Proof — step 2

- *V is normalized*
- For any $s \in A$,

$$
\begin{aligned}
& - V_A(x) \\
&= \sum_{B \subset A, s \notin B} (-1)^{|A-B|} \log \Pi(x^B) + \sum_{B \subset A, s \in B} (-1)^{|A-B|} \log \Pi(x^B) \\
&= \sum_{B \subset A - \{s\}} (-1)^{|A-B|} \log \Pi(x^B) \\
&\qquad + \sum_{B' \subset A - \{s\}} (-1)^{|A-B'-\{s\}|} \log \Pi(x^{B' \cup \{s\}}) \\
&= \sum_{B \subset A - \{s\}} (-1)^{|A-B|} \left(\log \Pi(x^B) - \log \Pi(x^{B \cup \{s\}}) \right).
\end{aligned}
$$

- If $x_s = 0$, then $x^B = x^{B \cup \{s\}}$.
- Hence $V_A(x) = 0$.

Proof — step 3 (1)

- *Proof of (48)*
- If $s \notin B$, and $B \subset A - \{s\}$,

$$x_{(s)}^{B} = x_{(s)}^{B\cup\{s\}}.$$

- Because

$$\Pi(x^B) = \Pi(x_s^B | x_{(s)}^B)\Pi(x_{(s)}^B)$$
$$\Pi(x^{B\cup\{s\}}) = \Pi(x_s^{B\cup\{s\}} | x_{(s)}^{B\cup\{s\}})\Pi(x_{(s)}^{B\cup\{s\}})$$

it follows that

$$\frac{\Pi(x^B)}{\Pi(x^{B\cup\{s\}})} = \frac{\Pi(x_s^B | x_{(s)}^B)}{\Pi(x_s^{B\cup\{s\}} | x_{(s)}^{B\cup\{s\}})}. \tag{52}$$

Proof — step 3 (1)

- *Proof of (48)*
- If $s \notin B$, and $B \subset A - \{s\}$,

$$x_{(s)}^{B} = x_{(s)}^{B \cup \{s\}}.$$

- Because

$$\Pi(x^B) = \Pi(x_s^B | x_{(s)}^B)\Pi(x_{(s)}^B)$$
$$\Pi(x^{B\cup\{s\}}) = \Pi(x_s^{B\cup\{s\}} | x_{(s)}^{B\cup\{s\}})\Pi(x_{(s)}^{B\cup\{s\}})$$

it follows that

$$\frac{\Pi(x^B)}{\Pi(x^{B\cup\{s\}})} = \frac{\Pi(x_s^B | x_{(s)}^B)}{\Pi(x_s^{B\cup\{s\}} | x_{(s)}^{B\cup\{s\}})}. \tag{52}$$

Proof — step 3 (1)

- *Proof of (48)*
- If $s \notin B$, and $B \subset A - \{s\}$,

$$x_{(s)}^{B} = x_{(s)}^{B \cup \{s\}}.$$

- Because

$$\Pi(x^B) = \Pi(x_s^B | x_{(s)}^B)\Pi(x_{(s)}^B)$$
$$\Pi(x^{B\cup\{s\}}) = \Pi(x_s^{B\cup\{s\}} | x_{(s)}^{B\cup\{s\}})\Pi(x_{(s)}^{B\cup\{s\}})$$

it follows that

$$\frac{\Pi(x^B)}{\Pi(x^{B\cup\{s\}})} = \frac{\Pi(x_s^B | x_{(s)}^B)}{\Pi(x_s^{B\cup\{s\}} | x_{(s)}^{B\cup\{s\}})}. \quad (52)$$

Proof — step 3 (2)

- *Proof of (48)*
- As in *step 2*

$$-V_A(x) = \sum_{B \subset A - \{s\}} (-1)^{|A-B|} \left(\log \Pi(x^B) - \log \Pi(x^{B \cup \{s\}}) \right)$$

$$= \sum_{B \subset A - \{s\}} (-1)^{|A-B|} \left(\log \Pi(x_s^B | x_{(s)}^B) - \log \Pi(x_s^{B \cup \{s\}} | x_{(s)}^{B \cup \{s\}}) \right)$$

The result follows by reversing the procedure in *step 2*.

Proof — step 3 (2)

- *Proof of (48)*
- As in *step 2*

$$-V_A(x) = \sum_{B \subset A - \{s\}} (-1)^{|A-B|} \left(\log \Pi(x^B) - \log \Pi(x^{B \cup \{s\}}) \right)$$

$$= \sum_{B \subset A - \{s\}} (-1)^{|A-B|} \left(\log \Pi(x_s^B | x_{(s)}^B) - \log \Pi(x_s^{B \cup \{s\}} | x_{(s)}^{B \cup \{s\}}) \right)$$

The result follows by reversing the procedure in *step 2*.

Proof — step 4 (1)

- *Uniqueness*.
- Assume U_A is another normalized potential such that

$$\Pi(x) = Z^{-1}\mathbf{e}^{-H_U(x)},$$

  where $H_U(x) = \sum_{B\subset S} U_B(x)$.
- By the normalization condition,

$$\Pi(0) = Z^{-1}\mathbf{e}^{-H_U(0)} = Z^{-1}.$$

- Hence

$$-H_U(x) = \log\left[\frac{\Pi(x)}{\Pi(0)}\right]. \quad (53)$$

- For any set $A \subset S$, $A \neq \emptyset$, let

$$\Phi(B) = -U_B(x^A), \quad \Psi(A) = \log\left[\frac{\Pi(x^A)}{\Pi(0)}\right].$$

- Then

$$\Psi(A) = -H_U(x^A) = -\sum_{B\subset S} U_B(x^A) = \sum_{B\subset A} \Phi(B).$$

Proof — step 4 (1)

- *Uniqueness*.
- Assume U_A is another normalized potential such that

$$\Pi(x) = Z^{-1}\mathbf{e}^{-H_U(x)},$$

where $H_U(x) = \sum_{B\subset S} U_B(x)$.

- By the normalization condition,

$$\Pi(0) = Z^{-1}\mathbf{e}^{-H_U(0)} = Z^{-1}.$$

- Hence

$$-H_U(x) = \log\left[\frac{\Pi(x)}{\Pi(0)}\right]. \tag{53}$$

- For any set $A \subset S$, $A \neq \emptyset$, let

$$\Phi(B) = -U_B(x^A), \quad \Psi(A) = \log\left[\frac{\Pi(x^A)}{\Pi(0)}\right].$$

- Then

$$\Psi(A) = -H_U(x^A) = -\sum_{B\subset S} U_B(x^A) = \sum_{B\subset A} \Phi(B).$$

Proof — step 4 (1)

- *Uniqueness*.
- Assume U_A is another normalized potential such that

$$\Pi(x) = Z^{-1}\mathbf{e}^{-H_U(x)},$$

where $H_U(x) = \sum_{B\subset S} U_B(x)$.
- By the normalization condition,

$$\Pi(0) = Z^{-1}\mathbf{e}^{-H_U(0)} = Z^{-1}.$$

- Hence

$$-H_U(x) = \log\left[\frac{\Pi(x)}{\Pi(0)}\right]. \tag{53}$$

- For any set $A \subset S$, $A \neq \emptyset$, let

$$\Phi(B) = -U_B(x^A), \quad \Psi(A) = \log\left[\frac{\Pi(x^A)}{\Pi(0)}\right].$$

- Then

$$\Psi(A) = -H_U(x^A) = -\sum_{B\subset S} U_B(x^A) = \sum_{B\subset A} \Phi(B).$$

Proof — step 4 (1)

- *Uniqueness*.
- Assume U_A is another normalized potential such that
$$\Pi(x) = Z^{-1}\mathbf{e}^{-H_U(x)},$$
where $H_U(x) = \sum_{B\subset S} U_B(x)$.
- By the normalization condition,
$$\Pi(0) = Z^{-1}\mathbf{e}^{-H_U(0)} = Z^{-1}.$$
- Hence
$$-H_U(x) = \log\left[\frac{\Pi(x)}{\Pi(0)}\right]. \tag{53}$$
- For any set $A \subset S$, $A \neq \emptyset$, let
$$\Phi(B) = -U_B(x^A), \quad \Psi(A) = \log\left[\frac{\Pi(x^A)}{\Pi(0)}\right].$$
- Then
$$\Psi(A) = -H_U(x^A) = -\sum_{B\subset S} U_B(x^A) = \sum_{B\subset A} \Phi(B).$$

Proof — step 4 (1)

- *Uniqueness*.
- Assume U_A is another normalized potential such that

$$\Pi(x) = Z^{-1}\mathbf{e}^{-H_U(x)},$$

  where $H_U(x) = \sum_{B\subset S} U_B(x)$.
- By the normalization condition,

$$\Pi(0) = Z^{-1}\mathbf{e}^{-H_U(0)} = Z^{-1}.$$

- Hence

$$-H_U(x) = \log\left[\frac{\Pi(x)}{\Pi(0)}\right]. \qquad (53)$$

- For any set $A \subset S$, $A \neq \emptyset$, let

$$\Phi(B) = -U_B(x^A), \quad \Psi(A) = \log\left[\frac{\Pi(x^A)}{\Pi(0)}\right].$$

- Then

$$\Psi(A) = -H_U(x^A) = -\sum_{B\subset S} U_B(x^A) = \sum_{B\subset A} \Phi(B).$$

Proof — step 4 (1)

- *Uniqueness*.
- Assume U_A is another normalized potential such that
$$\Pi(x) = Z^{-1}\mathbf{e}^{-H_U(x)},$$
where $H_U(x) = \sum_{B\subset S} U_B(x)$.
- By the normalization condition,
$$\Pi(0) = Z^{-1}\mathbf{e}^{-H_U(0)} = Z^{-1}.$$
- Hence
$$-H_U(x) = \log\left[\frac{\Pi(x)}{\Pi(0)}\right]. \tag{53}$$
- For any set $A \subset S$, $A \neq \emptyset$, let
$$\Phi(B) = -U_B(x^A), \quad \Psi(A) = \log\left[\frac{\Pi(x^A)}{\Pi(0)}\right].$$
- Then
$$\Psi(A) = -H_U(x^A) = -\sum_{B\subset S} U_B(x^A) = \sum_{B\subset A} \Phi(B).$$

Proof — step 4 (2)

- *Uniqueness*.
- By the Möbius inversion formula and the lemma

$$\Phi(A) = \sum_{B \subset A} (-1)^{|A-B|} \Psi(B)$$
$$= \sum_{B \subset A} (-1)^{|A-B|} \log \left[\frac{\Pi(x^A)}{\Pi(0)} \right]$$
$$= \sum_{B \subset A} (-1)^{|A-B|} \log \Pi(x^A),$$

- Because

$$\Phi(A) = -U_A(x^A) = -U_A(x), \quad \forall x \in \Omega$$

by (34), the proof is completed.

Proof — step 4 (2)

- *Uniqueness*.
- By the Möbius inversion formula and the lemma

$$\Phi(A) = \sum_{B \subset A} (-1)^{|A-B|} \Psi(B)$$
$$= \sum_{B \subset A} (-1)^{|A-B|} \log \left[\frac{\Pi(x^A)}{\Pi(0)} \right]$$
$$= \sum_{B \subset A} (-1)^{|A-B|} \log \Pi(x^A),$$

- Because

$$\Phi(A) = -U_A(x^A) = -U_A(x), \quad \forall x \in \Omega$$

by (34), the proof is completed.

Proof — step 4 (2)

- *Uniqueness*.
- By the Möbius inversion formula and the lemma

$$\begin{aligned}\Phi(A) &= \sum_{B\subset A}(-1)^{|A-B|}\Psi(B)\\ &= \sum_{B\subset A}(-1)^{|A-B|}\log\left[\frac{\Pi(x^A)}{\Pi(0)}\right]\\ &= \sum_{B\subset A}(-1)^{|A-B|}\log\Pi(x^A),\end{aligned}$$

- Because

$$\Phi(A) = -U_A(x^A) = -U_A(x), \quad \forall x \in \Omega$$

by (34), the proof is completed.

Outline

Markov Random Fields

Random Fields

Neighborhood Systems and Cliques

Markov Random Fields

Gibbsian Random Fields

Equivalence Theorem

Equivalence Theorem

Theorem

Let $\mathcal{G}$ be a neighborhood system on S. Then Π is a `Gibbsian random field` w.r.t $\mathcal{G}$ if and only if Π is a `Markov random field` w.r.t $\mathcal{G}$, in which case $\{V_A\}$ in (46) is a `Gibbsian potential`.

- The original version is in [Hammersley and Clifford, 1968] and others under some restrictions; see [Kinderman and Snell, 1980] and the references therein. The statement and proof here are essentially due to [Grimmett, 1973].

Equivalence Theorem

Theorem

Let $\mathcal{G}$ be a neighborhood system on S. Then Π is a `Gibbsian random field` *w.r.t $\mathcal{G}$ if and only if Π is a* `Markov random field` *w.r.t $\mathcal{G}$, in which case $\{V_A\}$ in (46) is a* `Gibbsian potential`.

- The original version is in [Hammersley and Clifford, 1968] and others under some restrictions; see [Kinderman and Snell, 1980] and the references therein. The statement and proof here are essentially due to [Grimmett, 1973].

Proof: " $\Longrightarrow$ " (1)

- Let Π have a Gibbsian representation w.r.t. $\mathcal{G}$ for some V:

$$\Pi(x) = Z^{-1}\mathbf{e}^{-H(x)}, \quad H(x) = \sum_{C\in\mathcal{C}} V_C(x). \tag{54}$$

- For $x \in \Omega$, $s \in S$, $\lambda \in \Lambda$, let $(\lambda, x_{(s)})$ denote the configuration obtained by replacing x_s by λ:

$$(\lambda, x_{(s)})_r = \begin{cases} x_r, & r \neq s, \\ \lambda, & r = s. \end{cases}$$

Proof: " $\Longrightarrow$ " (1)

- Let Π have a Gibbsian representation w.r.t. $\mathcal{G}$ for some V:

$$\Pi(x) = Z^{-1}\mathbf{e}^{-H(x)}, \quad H(x) = \sum_{C\in\mathcal{C}} V_C(x). \tag{54}$$

- For $x \in \Omega$, $s \in S$, $\lambda \in \Lambda$, let $(\lambda, x_{(s)})$ denote the configuration obtained by replacing x_s by λ:

$$(\lambda, x_{(s)})_r = \begin{cases} x_r, & r \neq s, \\ \lambda, & r = s. \end{cases}$$

Proof: " $\Longrightarrow$ " (2)

- Because $V_A(\lambda, x_{(s)}) = V_A(x)$ if $s \notin A$,

$$\Pi_s(x_s|x_{(s)}) = \frac{e^{-H(x)}}{\sum_{\lambda\in\Lambda} e^{-H(\lambda,x_{(s)})}}$$
$$= \frac{e^{-\sum_{A\in\mathcal{C},s\notin A} V_A(x) - \sum_{A\in\mathcal{C},s\in A} V_A(x)}}{\sum_{\lambda\in\Lambda} e^{-\sum_{A\in\mathcal{C},s\notin A} V_A(\lambda,x_{(s)}) - \sum_{A\in\mathcal{C},s\in A} V_A(\lambda,x_{(s)})}}$$
$$= \frac{e^{-\sum_{A\in\mathcal{C},s\in A} V_A(x)}}{\sum_{\lambda\in\Lambda} e^{-\sum_{A\in\mathcal{C},s\in A} V_A(\lambda,x_{(s)})}}.$$

- $A \in \mathcal{C}$ and $s \in A$ imply that $A \subset G_s \cup \{s\}$.
- Hence $\Pi_s(x_s|x_{(s)})$ depends only on x_t for $t \in G_s \cup \{s\}$, and it follows that,

$$\Pi_s(x_s|x_{(s)}) = \frac{e^{-\sum_{A\in\mathcal{C},s\in A\subset G_s\cup\{s\}} V_A(x)}}{\sum_{\lambda\in\Lambda_s} e^{-\sum_{A\in\mathcal{C},s\in A\subset G_s\cup\{s\}} V_A(\lambda,x_{(s)})}}.$$

Proof: " $\Longrightarrow$ " (2)

- Because $V_A(\lambda, x_{(s)}) = V_A(x)$ if $s \notin A$,

$$\Pi_s(x_s|x_{(s)}) = \frac{e^{-H(x)}}{\sum_{\lambda\in\Lambda} e^{-H(\lambda,x_{(s)})}}$$
$$= \frac{e^{-\sum_{A\in\mathcal{C},s\notin A} V_A(x) - \sum_{A\in\mathcal{C},s\in A} V_A(x)}}{\sum_{\lambda\in\Lambda} e^{-\sum_{A\in\mathcal{C},s\notin A} V_A(\lambda,x_{(s)}) - \sum_{A\in\mathcal{C},s\in A} V_A(\lambda,x_{(s)})}}$$
$$= \frac{e^{-\sum_{A\in\mathcal{C},s\in A} V_A(x)}}{\sum_{\lambda\in\Lambda} e^{-\sum_{A\in\mathcal{C},s\in A} V_A(\lambda,x_{(s)})}}.$$

- $A \in \mathcal{C}$ and $s \in A$ imply that $A \subset G_s \cup \{s\}$.
- Hence $\Pi_s(x_s|x_{(s)})$ depends only on x_t for $t \in G_s \cup \{s\}$, and it follows that,

$$\Pi_s(x_s|x_{(s)}) = \frac{e^{-\sum_{A\in\mathcal{C},s\in A\subset G_s\cup\{s\}} V_A(x)}}{\sum_{\lambda\in\Lambda_s} e^{-\sum_{A\in\mathcal{C},s\in A\subset G_s\cup\{s\}} V_A(\lambda,x_{(s)})}}.$$

Proof: " $\Longrightarrow$ " (2)

- Because $V_A(\lambda, x_{(s)}) = V_A(x)$ if $s \notin A$,

$$\Pi_s(x_s|x_{(s)}) = \frac{e^{-H(x)}}{\sum_{\lambda\in\Lambda} e^{-H(\lambda,x_{(s)})}}$$
$$= \frac{e^{-\sum_{A\in\mathcal{C},s\notin A} V_A(x) - \sum_{A\in\mathcal{C},s\in A} V_A(x)}}{\sum_{\lambda\in\Lambda} e^{-\sum_{A\in\mathcal{C},s\notin A} V_A(\lambda,x_{(s)}) - \sum_{A\in\mathcal{C},s\in A} V_A(\lambda,x_{(s)})}}$$
$$= \frac{e^{-\sum_{A\in\mathcal{C},s\in A} V_A(x)}}{\sum_{\lambda\in\Lambda} e^{-\sum_{A\in\mathcal{C},s\in A} V_A(\lambda,x_{(s)})}}.$$

- $A \in \mathcal{C}$ and $s \in A$ imply that $A \subset G_s \cup \{s\}$.
- Hence $\Pi_s(x_s|x_{(s)})$ depends only on x_t for $t \in G_s \cup \{s\}$, and it follows that,

$$\Pi_s(x_s|x_{(s)}) = \frac{e^{-\sum_{A\in\mathcal{C},s\in A\subset G_s\cup\{s\}} V_A(x)}}{\sum_{\lambda\in\Lambda_s} e^{-\sum_{A\in\mathcal{C},s\in A\subset G_s\cup\{s\}} V_A(\lambda,x_{(s)})}}.$$

Proof: " $\Longrightarrow$ " (3)

- For the derivation of $\Pi(x_s | x_r, r \in G_s)$, introduce the following notations.
- Let
$$J = \{j : j \notin G_s \cup \{s\}\}.$$
- For $x \in \Omega$, $\lambda \in \Lambda_s$, $\lambda_j \in \Lambda_j$, $j \in J$, let $(\lambda_J, x_{(J)})$ denote the configuration
$$(\lambda_J, x_{(J)})_r = \begin{cases} x_r, & r \in G_s \cup \{s\}, \\ \lambda_j, & r \notin G_s \cup \{s\}, \end{cases}$$
- let $(\lambda_J, x_{(J+s)})$ denote the configuration such that
$$(\lambda_J, x_{(J+s)})_r = \begin{cases} x_r, & r \in G_s, \\ \lambda, & r = s, \\ \lambda_j, & r \notin G_s. \end{cases}$$

Proof: " $\Longrightarrow$ " (3)

- For the derivation of $\Pi(x_s | x_r, r \in G_s)$, introduce the following notations.
- Let
$$J = \{j : j \notin G_s \cup \{s\}\}.$$
- For $x \in \Omega$, $\lambda \in \Lambda_s$, $\lambda_j \in \Lambda_j$, $j \in J$, let $(\lambda_J, x_{(J)})$ denote the configuration
$$(\lambda_J, x_{(J)})_r = \begin{cases} x_r, & r \in G_s \cup \{s\}, \\ \lambda_j, & r \notin G_s \cup \{s\}, \end{cases}$$
- let $(\lambda_J, x_{(J+s)})$ denote the configuration such that
$$(\lambda_J, x_{(J+s)})_r = \begin{cases} x_r, & r \in G_s, \\ \lambda, & r = s, \\ \lambda_j, & r \notin G_s. \end{cases}$$

Proof: " $\Longrightarrow$ " (3)

- For the derivation of $\Pi(x_s | x_r, r \in G_s)$, introduce the following notations.
- Let
$$J = \{j : j \notin G_s \cup \{s\}\}.$$
- For $x \in \Omega$, $\lambda \in \Lambda_s$, $\lambda_j \in \Lambda_j$, $j \in J$, let $(\lambda_J, x_{(J)})$ denote the configuration
$$(\lambda_J, x_{(J)})_r = \begin{cases} x_r, & r \in G_s \cup \{s\}, \\ \lambda_j, & r \notin G_s \cup \{s\}, \end{cases}$$
- let $(\lambda_J, x_{(J+s)})$ denote the configuration such that
$$(\lambda_J, x_{(J+s)})_r = \begin{cases} x_r, & r \in G_s, \\ \lambda, & r = s, \\ \lambda_j, & r \notin G_s. \end{cases}$$

Proof: " $\Longrightarrow$ " (3)

- For the derivation of $\Pi(x_s | x_r, r \in G_s)$, introduce the following notations.
- Let
$$J = \{j : j \notin G_s \cup \{s\}\}.$$
- For $x \in \Omega$, $\lambda \in \Lambda_s$, $\lambda_j \in \Lambda_j$, $j \in J$, let $(\lambda_J, x_{(J)})$ denote the configuration
$$(\lambda_J, x_{(J)})_r = \begin{cases} x_r, & r \in G_s \cup \{s\}, \\ \lambda_j, & r \notin G_s \cup \{s\}, \end{cases}$$
- let $(\lambda_J, x_{(J+s)})$ denote the configuration such that
$$(\lambda_J, x_{(J+s)})_r = \begin{cases} x_r, & r \in G_s, \\ \lambda, & r = s, \\ \lambda_j, & r \notin G_s. \end{cases}$$

Proof: " $\Longrightarrow$ " (4)

- Then

$$\Pi(x_s | x_r, r \in G_s)$$

$$= \frac{\Pi(x_s, x_r, r \in G_s)}{\Pi(x_r, r \in G_s)}$$

$$= \frac{\sum_{\substack{\lambda_j \in \Lambda_j \\ j \in J}} \Pi(\lambda_j, j \in J, x_s, x_r, r \in G_s)}{\sum_{\substack{\lambda_j \in \Lambda_j \\ j \in J}} \sum_{\lambda \in \Lambda_s} \Pi(\lambda_j, j \in J, \lambda, x_r, r \in G_s)}$$

$$= \frac{\sum_{\substack{\lambda_j \in \Lambda_j \\ j \in J}} \mathbf{e}^{-H(\lambda_J, x_{(J)})}}{\sum_{\substack{\lambda_j \in \Lambda_j \\ j \in J}} \sum_{\lambda \in \Lambda_s} \mathbf{e}^{-H(\lambda_J, x_{(J+s)})}}$$

$$= \frac{\sum_{\substack{\lambda_j \in \Lambda_j \\ j \in J}} \mathbf{e}^{-\sum_{A \in \mathcal{C}, s \notin A} V_A(\lambda_J, x_{(J)}) - \sum_{A \in \mathcal{C}, s \in A} V_A(\lambda_J, x_{(J)})}}{\sum_{\lambda \in \Lambda_s} \sum_{\substack{\lambda_j \in \Lambda_j \\ j \in J}} \mathbf{e}^{-\sum_{A \in \mathcal{C}, s \notin A} V_A(\lambda_J, x_{(J+s)}) - \sum_{A \in \mathcal{C}, s \in A} V_A(\lambda_J, x_{(J+s)})}}$$

Proof: $'' \Longrightarrow''$ (5)

- For every $A \in \mathcal{C}$, if $s \notin A$,

$$V_A(\lambda_J, x_{(J)}) = V_A(\lambda_J, x_{(J+s)}).$$

- Therefore

$$\Pi(x_s | x_r, r \in G_s) = \frac{\sum_{\substack{\lambda_j \in \Lambda_j \\ j \in J}} \mathbf{e}^{-\sum_{A \in \mathcal{C}, s \in A \subset G_s \cup \{s\}} V_A(\lambda_J, x_{(J)})}}{\sum_{\lambda \in \Lambda_s} \sum_{\substack{\lambda_j \in \Lambda_j \\ j \in J}} \mathbf{e}^{-\sum_{A \in \mathcal{C}, s \in A \subset G_s \cup \{s\}} V_A(\lambda_J, x_{(J+s)})}}.$$

- For $s \in A \subset G_s \cup \{s\}$,

$$V_A(\lambda_J, x_{(J)}) = V_A(x), \tag{55}$$

$$V_A(\lambda_J, x_{(J+s)}) = V_A(\lambda, x_{(s)}). \tag{56}$$

Proof: " $\Longrightarrow$ " (5)

- For every $A \in \mathcal{C}$, if $s \notin A$,

$$V_A(\lambda_J, x_{(J)}) = V_A(\lambda_J, x_{(J+s)}).$$

- Therefore

$$\Pi(x_s | x_r, r \in G_s) = \frac{\sum_{\substack{\lambda_j \in \Lambda_j \\ j \in J}} \mathbf{e}^{-\sum_{A \in \mathcal{C}, s \in A \subset G_s \cup \{s\}} V_A(\lambda_J, x_{(J)})}}{\sum_{\lambda \in \Lambda_s} \sum_{\substack{\lambda_j \in \Lambda_j \\ j \in J}} \mathbf{e}^{-\sum_{A \in \mathcal{C}, s \in A \subset G_s \cup \{s\}} V_A(\lambda_J, x_{(J+s)})}}.$$

- For $s \in A \subset G_s \cup \{s\}$,

$$V_A(\lambda_J, x_{(J)}) = V_A(x), \quad (55)$$

$$V_A(\lambda_J, x_{(J+s)}) = V_A(\lambda, x_{(s)}). \quad (56)$$

Proof: $'' \Longrightarrow''$ (5)

- For every $A \in \mathcal{C}$, if $s \notin A$,

$$V_A(\lambda_J, x_{(J)}) = V_A(\lambda_J, x_{(J+s)}).$$

- Therefore

$$\Pi(x_s | x_r, r \in G_s) = \frac{\sum_{\substack{\lambda_j \in \Lambda_j \\ j \in J}} \mathbf{e}^{-\sum_{A \in \mathcal{C}, s \in A \subset G_s \cup \{s\}} V_A(\lambda_J, x_{(J)})}}{\sum_{\lambda \in \Lambda_s} \sum_{\substack{\lambda_j \in \Lambda_j \\ j \in J}} \mathbf{e}^{-\sum_{A \in \mathcal{C}, s \in A \subset G_s \cup \{s\}} V_A(\lambda_J, x_{(J+s)})}}.$$

- For $s \in A \subset G_s \cup \{s\}$,

$$V_A(\lambda_J, x_{(J)}) = V_A(x), \tag{55}$$

$$V_A(\lambda_J, x_{(J+s)}) = V_A(\lambda, x_{(s)}). \tag{56}$$

Proof: " $\Longrightarrow$ " (6)

- Hence

$$\Pi(x_s|x_r, r \in G_s) = \frac{\sum_{\substack{\lambda_j \in \Lambda_j \\ j \in J}} \mathbf{e}^{-\sum_{A \in \mathcal{C}, s \in A \subset G_s \cup \{s\}} V_A(x)}}{\sum_{\lambda \in \Lambda_s} \sum_{\substack{\lambda_j \in \Lambda_j \\ j \in J}} \mathbf{e}^{-\sum_{A \in \mathcal{C}, s \in A \subset G_s \cup \{s\}} V_A(\lambda, x_{(s)})}} \tag{57}$$

$$= \Pi_s(x_s|x_{(s)}). \tag{58}$$

Proof: " $\Longleftarrow$ " (1)

- Now suppose Π is a MRF w.r.t. G and let $V = (V_A)$ be the canonical potential associated with Π as in (46) or (48).
- The proof will be completed by showing that $V_A(x) = 0$ if $A \notin \mathcal{C}$.
- Choose $A \notin \mathcal{C}$.
- There $\exists s, t \in A$ such that $t \notin G_s \cup \{s\}$.

Proof: $'' \Longleftarrow ''$ (1)

- Now suppose Π is a MRF w.r.t. G and let $V = (V_A)$ be the canonical potential associated with Π as in (46) or (48).
- The proof will be completed by showing that $V_A(x) = 0$ if $A \notin \mathcal{C}$.
- Choose $A \notin \mathcal{C}$.
- There $\exists s, t \in A$ such that $t \notin G_s \cup \{s\}$.

Proof: " $\Longleftarrow$ " (1)

- Now suppose Π is a MRF w.r.t. G and let $V = (V_A)$ be the canonical potential associated with Π as in (46) or (48).
- The proof will be completed by showing that $V_A(x) = 0$ if $A \notin \mathcal{C}$.
- Choose $A \notin \mathcal{C}$.
- There $\exists s, t \in A$ such that $t \notin G_s \cup \{s\}$.

Proof: " $\Longleftarrow$ " (1)

- Now suppose Π is a MRF w.r.t. G and let $V = (V_A)$ be the canonical potential associated with Π as in (46) or (48).
- The proof will be completed by showing that $V_A(x) = 0$ if $A \notin \mathcal{C}$.
- Choose $A \notin \mathcal{C}$.
- There $\exists s, t \in A$ such that $t \notin G_s \cup \{s\}$.

Proof: $'' \Longleftarrow ''$ (2)

$$
\begin{aligned}
-V_A(x) &= \sum_{B\subset A} (-1)^{|A-B|} \log \Pi(x_s^B | x_{(s)}^B) \\
&= \sum_{B\subset A, s\notin B, t\notin B} (-1)^{|A-B|} \log \Pi(x_s^B | x_{(s)}^B) + \sum_{B\subset A, s\in B, t\notin B} (-1)^{|A-B|} \log \Pi(x_s^B | x_{(s)}^B) \\
&\quad + \sum_{B\subset A, s\notin B, t\in B} (-1)^{|A-B|} \log \Pi(x_s^B | x_{(s)}^B) + \sum_{B\subset A, s\in B, t\in B} (-1)^{|A-B|} \log \Pi(x_s^B | x_{(s)}^B) \\
&= \sum_{B\subset A-\{s\}-\{t\}} (-1)^{|A-B|} \log \Pi(x_s^B | x_{(s)}^B) + \sum_{s\in B_1\subset A-\{t\}} (-1)^{|A-B_1|} \log \Pi(x_s^{B_1} | x_{(s)}^{B_1}) \\
&\quad + \sum_{t\in B_2\subset A-\{s\}} (-1)^{|A-B_2|} \log \Pi(x_s^{B_2} | x_{(s)}^{B_2}) + \sum_{\{s,t\}\subset B_3\subset A} (-1)^{|A-B_3|} \log \Pi(x_s^{B_3} | x_{(s)}^{B_3}) \\
&= \sum_{B\subset A-\{s\}-\{t\}} (-1)^{|A-B|} \log \Pi(x_s^B | x_{(s)}^B) \\
&\quad + \sum_{B\subset A-\{s\}-\{t\}} (-1)^{|A-B-\{s\}|} \log \Pi(x_s^{B\cup\{s\}} | x_{(s)}^{B\cup\{s\}}) \\
&\quad + \sum_{B\subset A-\{s\}-\{t\}} (-1)^{|A-B-\{t\}|} \log \Pi(x_s^{B\cup\{t\}} | x_{(s)}^{B\cup\{t\}}) \\
&\quad + \sum_{B\subset A-\{s\}-\{t\}} (-1)^{|A-B-\{s\}-\{t\}|} \log \Pi(x_s^{B\cup\{s,t\}} | x_{(s)}^{B\cup\{s,t\}}).
\end{aligned}
$$

Proof: "$\Longleftarrow$" (3)

$$V_A(x) = \sum_{B \subset A - \{s\} - \{t\}} (-1)^{|A-B|} \log \left[\frac{\Pi(x_s^B | x_{(s)}^B)\Pi(x_s^{B\cup\{s,t\}} | x_{(s)}^{B\cup\{s,t\}})}{\Pi(x_s^{B\cup\{s\}} | x_{(s)}^{B\cup\{s\}})\Pi(x_s^{B\cup\{t\}} | x_{(s)}^{B\cup\{t\}})} \right].$$

- By the MRF property,

$$\Pi(x_s^B | x_{(s)}^B) = \Pi(x_s^B | x_r^B, r \in G_s),$$
$$\Pi(x_s^{B\cup\{t\}} | x_{(s)}^{B\cup\{t\}}) = \Pi(x_s^{B\cup\{t\}} | x_r^{B\cup\{t\}}, r \in G_s).$$

- For every subset B of S, if $t \notin B$ and $r \neq t$, we have

$$x_r^B = x_r^{B\cup\{t\}}.$$

- Because $t \notin G_s \cup \{s\}$,

$$\Pi(x_s^B | x_{(s)}^B) = \Pi(x_s^{B\cup\{t\}} | x_{(s)}^{B\cup\{t\}}).$$

Proof: $'' \Longleftarrow ''$ (3)

$$V_A(x) = \sum_{B \subset A - \{s\} - \{t\}} (-1)^{|A-B|} \log \left[\frac{\Pi(x_s^B | x_{(s)}^B) \Pi(x_s^{B\cup\{s,t\}} | x_{(s)}^{B\cup\{s,t\}})}{\Pi(x_s^{B\cup\{s\}} | x_{(s)}^{B\cup\{s\}}) \Pi(x_s^{B\cup\{t\}} | x_{(s)}^{B\cup\{t\}})} \right].$$

- By the MRF property,

$$\Pi(x_s^B | x_{(s)}^B) = \Pi(x_s^B | x_r^B, r \in G_s),$$
$$\Pi(x_s^{B\cup\{t\}} | x_{(s)}^{B\cup\{t\}}) = \Pi(x_s^{B\cup\{t\}} | x_r^{B\cup\{t\}}, r \in G_s).$$

- For every subset B of S, if $t \notin B$ and $r \neq t$, we have

$$x_r^B = x_r^{B\cup\{t\}}.$$

- Because $t \notin G_s \cup \{s\}$,

$$\Pi(x_s^B | x_{(s)}^B) = \Pi(x_s^{B\cup\{t\}} | x_{(s)}^{B\cup\{t\}}).$$

Proof: " $\Longleftarrow$ " (3)

$$V_A(x) = \sum_{B \subset A - \{s\} - \{t\}} (-1)^{|A-B|} \log \left[\frac{\Pi(x_s^B | x_{(s)}^B) \Pi(x_s^{B \cup \{s,t\}} | x_{(s)}^{B \cup \{s,t\}})}{\Pi(x_s^{B \cup \{s\}} | x_{(s)}^{B \cup \{s\}}) \Pi(x_s^{B \cup \{t\}} | x_{(s)}^{B \cup \{t\}})} \right].$$

- By the MRF property,

$$\Pi(x_s^B | x_{(s)}^B) = \Pi(x_s^B | x_r^B, r \in G_s),$$
$$\Pi(x_s^{B \cup \{t\}} | x_{(s)}^{B \cup \{t\}}) = \Pi(x_s^{B \cup \{t\}} | x_r^{B \cup \{t\}}, r \in G_s).$$

- For every subset B of S, if $t \notin B$ and $r \neq t$, we have

$$x_r^B = x_r^{B \cup \{t\}}.$$

- Because $t \notin G_s \cup \{s\}$,

$$\Pi(x_s^B | x_{(s)}^B) = \Pi(x_s^{B \cup \{t\}} | x_{(s)}^{B \cup \{t\}}).$$

Proof: " $\Longleftarrow$ " (4)

- Similarly, we have

$$\Pi(x_s^{B\cup\{s\}}|x_{(s)}^{B\cup\{s\}}) = \Pi(x_s^{B\cup\{s,t\}}|x_{(s)}^{B\cup\{s,t\}})$$

and consequently that $V_A(x) = 0$.

Remark

If V is a Gibbsian potential, we have seen in the above theorem that

$$\Pi(x_s|x_{(s)}) = Z_s^{-1}\mathbf{e}^{-\sum_{A\in\mathcal{C},s\in A} V_A(x)} \quad (59)$$

$$Z_s = \sum_{\lambda\in\Lambda_s} \mathbf{e}^{-\sum_{A\in\mathcal{C},s\in A} V_A(\lambda,x_{(s)})} \quad (60)$$

Proof: $'' \Longleftarrow ''$ (4)

- Similarly, we have

$$\Pi(x_s^{B\cup\{s\}}|x_{(s)}^{B\cup\{s\}}) = \Pi(x_s^{B\cup\{s,t\}}|x_{(s)}^{B\cup\{s,t\}})$$

and consequently that $V_A(x) = 0$.

Remark

If V is a Gibbsian potential, we have seen in the above theorem that

$$\Pi(x_s|x_{(s)}) = Z_s^{-1}\mathbf{e}^{-\sum_{A\in\mathcal{C},s\in A} V_A(x)} \tag{59}$$

$$Z_s = \sum_{\lambda\in\Lambda_s} \mathbf{e}^{-\sum_{A\in\mathcal{C},s\in A} V_A(\lambda,x_{(s)})} \tag{60}$$

References I

Geman, D. (1990).
Random fields and inverse problems in imaging.
In Ancona, A., Geman, D., Ikeda, N., and Geman, D., editors, *École d'Été de Probabilités de Saint-Flour XVIII i– 1988*, volume 1427 of *Lecture Notes in Mathematics*, pages 113–193. Springer Berlin / Heidelberg.

Geman, S. and Geman, D. (1984).
Stochastic relaxation, gibbs distributions and the bayesian restoration of images.
IEEE Transactions on Pattern Analysis and Machine Intelligence, 6:721 – 741.

References II

- Geman, S. and Graffigne, C. (1987).
  Markov random field image models and their applications to computer vision.
  In Gleason, A. M., editor, *Proceedings of the International Congress of Mathematicians 1986*, pages 1496 – 1517, Providence, R.I. American Mathematical Society.
- Grimmett, G. R. (1973).
  A theorem about random fields.
  Bulletin of the London Mathematical Society, 5:81 – 84.
- Hammersley, J. M. and Clifford, P. (1968).
  Markov fields of finite graphs and lattices.
  Preprint, University of California, Berkeley.

References III

- Kinderman, R. and Snell, J. L. (1980).
  Markov Random Fields and Their Applications.
  American Mathematical Society, Providence, RI.
- Li, S. Z. (2009).
  Markov Random Field Modeling in Image Analysis.
  Springer Publishing Company, Incorporated, 3rd edition.
- Pérez, P. and Heitz, F. (1994).
  Restriction of a markov random field on a graph and multiresolution statistical image modeling.
  Technical Report 784, Institut de Recherche en Informatique et Systèmes Aléatoires, Campus de Beaulieu, Rennes Cedex, France.

References IV

- Rudin, W. (1970).
  Real and Complex Analysis.
  McGraw-Hill, New York.
- Wrinkler, G. (1995).
  Image Analysis, Random Fields and Dynamic Monte Carlo Methods.
  Springer, Berlin–Heideberg.