# Probabilistic Graphical Models

IKHAN CHOI

Contents



## 1. Introduction

### 1.1. Statistical model.

**Definition 1.1.** A *statistical model* is an approximation scheme for unknown joint probability distribution.

The general purpose of many statistical models is to *estimate the joint probability distribution* of several random variables. The joint probability distrbution contains data about relations of the random variables. For example, suppose our goal is to obtain the most possible value of $Y$ when given $X = x$, and we have already estimated the joint distribution function $f_{X,Y}$. Then, since the function $y \mapsto f_{X,Y}(x, y)$ describes the distribution of the random variable $Y|X = x$, what we want to find can be defined reasonably as

$$\widehat{y} = \arg\max_y f_{X,Y}(x, y).$$

**Example 1.1.** A random field, which we have not defined yet, is a way to represent several random variables together with their dependecies. Therefore, a paramterized random field gives a statistical model. In this case, training means an approximating process to find the best parameter, usually written as $\beta$.

### 1.2. Random fields.

**Definition 1.2** (Random field)**.** A *random field* is a set of random variables parametrized by a topological space or a (directed or undirected) graph.

**Definition 1.3.** In this note, a term *random field* or *network* will be used to refer to random fields on a graph.

*Last Update*: August 13, 2019.

**Example 1.2** (Markov chain)**.** Define a graph $G = (V, E)$ and $\mathcal{S}$ such that:

$$\begin{aligned} V &= \mathbb{Z}_{t\geq 0}, \\ E &= \{(t, t+1)\}_{t\in V}, \\ \mathcal{S} &= \text{a finite set.} \end{aligned}$$

An element $t \in V$ denotes the time $t$. Then, the set of $\mathcal{S}$-valued random variables $\{X_t\}_{t\in V}$ indexed by $V$ defines a random field.

The Markov property is given by

$$X_t \perp\!\!\!\perp X_s \mid X_{t-1}$$

for $s \leq t$. Since $\mathcal{S}$ is finite, alternatively we may rewrite it by

$$\Pr(X_t = x_t \mid X_{t-1} = x_{t-1}) = \Pr(X_t = x_t \mid X_{t-1} = x_{t-1}, \cdots, X_0 = x_0).$$

**Example 1.3** (Maxwell-Boltzman distribution)**.** Let $X$ be a $\mathcal{S}$-valued random field on a graph $G = (V, E)$ such that:

$$\begin{aligned} V &= \{n\}_{n=1}^{N}, \\ E &= \varnothing, \\ \mathcal{S} &= \mathbb{Z}_x^3 \times \mathbb{Z}_p^3 \end{aligned}$$

for a large natural number $N$. The set $V$ is the set of ideal gas particles, and the space of states $\mathcal{S}$ embodies the discretized phase space. At each particle $n \in V$ is attached an $\mathcal{S}$-valued random variable denoted by $X_n$. Our ultimate goal is to describe the joint probability distribution of $X_n$'s; equivalently, the distribution of a $\mathcal{S}^V$-valued random variable $X = (X_1, \cdots, X_N)$. Elements of $\mathcal{S}^V$ are called *microstates*.

Let $m > 0$ be a constant and fix a parameter $\beta = \frac{1}{k_B T} > 0$ which is called *coldness*. Define the *Boltzmann factor* as a function $\phi_n : \mathbb{R}_{\beta>0} \times \mathcal{S} \to \mathbb{R}$ such that

$$\phi_n(\beta, i) := e^{-\beta E(i)},$$

where the *energy function* $E : \mathcal{S} \to \mathbb{R}$ is defined by

$$E(x, p) := \frac{\|p\|^2}{2m}.$$

The assumption for Boltzmann factors states that given $\beta$, the probability for a particle to be in the state $i$ is proportional to the Boltzmann factor:

$$\Pr(X_n = i_n) \propto \phi_n(\beta, i_n) = e^{-\beta E_{i_n}}$$

for each state $i_n \in \mathcal{S}$ and a fixed particle $n \in V$. Thus, we can write

$$\Pr(X_n = i_n) = \frac{\phi_n(\beta, i_n)}{\sum_{j_n\in\mathcal{S}} \phi_n(\beta, j_n)} =: \frac{\phi_n(\beta, i_n)}{Z_n(\beta)}.$$

If we assume the independence of $X_n$'s, then we get the disjoint probability distribution

$$\Pr(X = i) = \frac{\phi(\beta, i)}{\sum_{j\in\mathcal{S}^V} \phi(\beta, j)} =: \frac{\phi(\beta, i)}{Z(\beta)},$$

where
$$\phi(\beta, i) := \prod_{n=1}^{N} \phi_n(\beta, i_n), \qquad Z(\beta) = \prod_{n=1}^{N} Z_n(\beta).$$
The denominator $Z : \mathbb{R}_{\beta>0} \to \mathbb{R}^+$ is called the *partition function.*

**Example 1.4** (Ising model)**.** Let $X$ be a $\mathcal{S}$-valued random field on a graph $G = (V, E)$ such that:
$$\begin{aligned} V &= \mathbb{Z}^2, \\ E &= \{\, (x, y) : \|x - y\| = 1 \,\}, \\ \mathcal{S} &= \{\pm 1\}. \end{aligned}$$

1.3. **Independencies.** We review the measure-theoretic definition of probability distribution.

**Definition 1.4.** Let $\Omega$ be a measure space with probability measure Pr and $\mathcal{S}$ a measurable space. Let $X : \Omega \to \mathcal{S}$ be a random variable. The *probability distribution* of $X$ is the pushforward measure $X_* \Pr$ on $\mathcal{S}$ defined as $X_* \Pr(A) = \Pr(X^{-1}(A))$ for measurable $A \in \mathcal{S}$. We often also write $\Pr(X \in A)$ for $X_* \Pr(A)$.

We mainly deal with several random variables and their joint distribution. Suppose $X = \{X_n\}_{n=1}^N$ is a set of random variables and let $(\mathcal{S}_n, \mathcal{F}_n)$ be the codomain of $X_n$.

**Definition 1.5.** Radon space?
$$\Pr(X \in A \mid Y \in B) := \lim_{U \downarrow B} \frac{\Pr(X \in A,\, Y \in U)}{\Pr(Y \in U)}.$$

Consequently, we may find
- $\Pr(X_1)$ is a probability measure on $\mathcal{S}_1$,
- $\Pr(X_1, X_2)$ is a probability measure on $\mathcal{S}_1 \times \mathcal{S}_2$,
- $\Pr(X_2 \mid X_1)$ is a set of probability measures on $\mathcal{S}_2$ parametrized by $\mathcal{F}_1$.

## 2. Bayesian networks

**Definition 2.1** (Bayesian network)**.** Let $G$ be a directed acyclic graph.

The graph acts as a parameter space. We want to investigate mutual effects among the paramtrized random variables.

**Theorem 2.1** (Factorization of probablity)**.**

**Example 2.2** (NBC, Naive Bayesian Classifier)**.**

**Example 2.3** (HMM, Hidden Markov Model)**.**

## 3. Markov networks

**Definition 3.1** (Markov network)**.**

Markov networks are sometimes called MRF, Markov random field.

**Example 3.1** (CRF, Conditional Random Field)**.** Consider a network with a graph $G$ such that vertices are divided into two classes.

## 4. Neural networks

Probabilistic graphical models provide effective explanations of the neural networks, but neural networks are not confined only to graphical models.

**Definition 4.1** (Neural network)**.** *Neural network* cannot be defined mathematically. It indicates statistical models that can solve problems with a collection of artificial neurons by adjusting connection strength among them.

**Example 4.1** (MLP, Multi-layer Perceptron)**.**

**Example 4.2** (RNN, Recurrent Neural Network)**.**