- Bayesian Philosophy
- Markov Random Fields
  - Introduction
  - Hammersley-Clifford Theorem
  - Dual Lattice Number

In last lecture, we discussed **Maximum Likelihood Estimation** and its application in **Mixture Model**. The core of Maximum Likelihood Estimation is to find fixed but unknown parameter θ that maximizes the likelihood function $P(y|\theta)$, Where y is our observation. Today, we begin to discuss another statistic concept that has broad applications in Computer Vision, **Bayesian Estimation**.

1 Bayesian Philosophy

Formal definition: in the Bayesian approach it is assumed that we have some probability density function for parameter θ. Then, the given the obsevations y the conditional probability density for that parameter is given by Eq. 1

$$P(\theta|y) = \frac{P(y|\theta)P(\theta)}{P(y)} \tag{1}$$

where $P(\theta)$ is *prior* probability, which is get before receiving any data. $P(\theta|y)$ is called *posterior* probability, which is obtained after those data coming. The computer vision analogy is given as

$$P(World\,|\,Image) = \frac{P(Image\,|\,World)P(World)}{P(Image)}$$

where $P(World)$ is our instinct guess with our eyes shut. For instance, we flip a coin n trials and observe y times of heads. $P(\theta)$ is the (prior) probability of seeing heads. If $P(\theta)$ is specified to be uniformly distributed over interval $[0, 1]$, the postperior probability $P(\theta|y)$ is obtained as

$$P(\theta|y) = \left[\binom{n}{y} P(\theta)^y (1 - P(\theta))^{n-y}\right] / z \tag{2}$$

where z is normalizing constant to make $P(\theta\,|\,y)$ within $[0, 1]$. With different sets of n and y, we have Figure 1. We see the conditional distribution function becomes more steeper

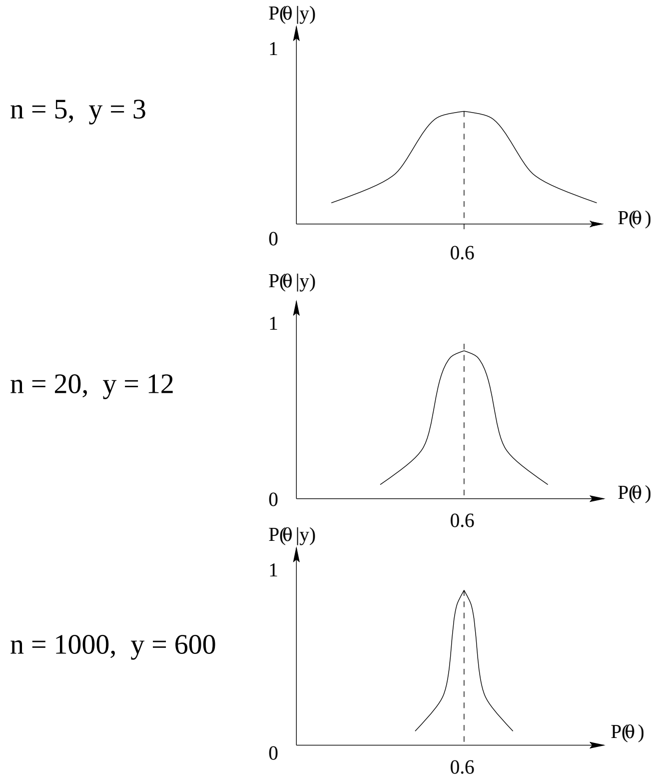

Figure 1: Coin flipping examples for Bayesian estimation.

as the number of trials rises up. This example demonstrate that how conditional probability (Bayesian estimation) can help us to determine objects in vision.

Suppose that we have 20 trials, then what is the probability of ≥ 15 heads show up?

$$P(\geq 15 heads) = \int_{x=15}^{20} \int_{\theta} P(x|\theta) dx P(\theta|y) d\theta$$

In vision, we generally assume that we have piecewise constant / slowly varying cues in our images. The cues can be pixel brightness, texture configuration and even motion. As in Figure 2, we see 5 different texture regions. Within each region, the texture structure maintains the same. In the real world, what we will observe is the real clean image I and additional noise η. Our next target is trying to eliminate the effect of η.

2 Markov Random Fields

2.1 Introduction

Markove Random Field (MRF) is introduced by Dobrushin in 1968 and refined by Stuart Geman and Donald Geman in 1984. This concept is an extension of **Markov Chain**. In an 1-D example, we put a drunk person on a line. At each time step he randomly walks to either the left or the right. As we show in Figure 3, if he reaches position n at time t, then

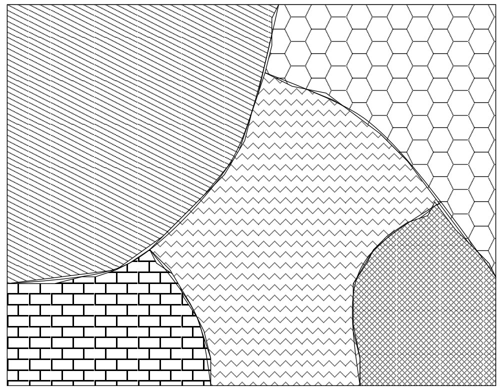

Figure 2: One image contains five different texture regions. The texture structure remains the same inside each region.

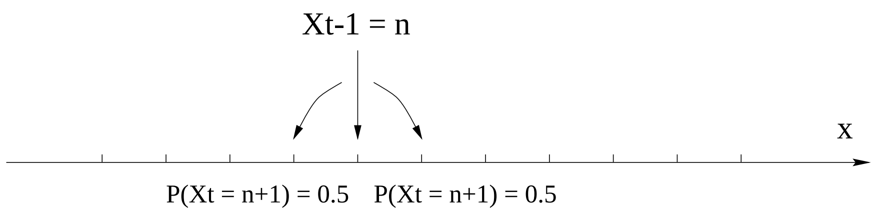

Figure 3: Drunk person walks on a line. An illustration of Markov Chain

probabilities of reaching position $n+1$ or $n-1$ are the same, 0.5. The probability that the person stay on position x at time t is given by

$$P(x_t|x_{t-1}, x_{t-2}, \ldots) = P(x_t|x_{t-1}) \tag{3}$$

We state that *"The future is independent of the past given the present"*.

The definition of MRF consists of 2 properties:

1. $P(X = \omega) > 0$, $\forall \omega \in \Omega$ where ω is a configuration as in the right half of Figure 4.
2. $P(X_s = x_s|X_r = x_r, r \neq s) = P(X_s = x_s|X_r = x_r, r \in G_s)$ where G_s is a neighborhood of s.

For example, $P(x_4 \,|\, x_1,\, x_2,\, x_3,\, x_5,\, \ldots,\, x_9) = P(x_4 \,|\, x_1,\, x_5,\, x_7)$. MRF is also called undirected graphical model.

2.2 Hammersley-Clifford Theorem

Deifintion of Hammersley-Clifford Theorem is given by

$$P(x = \omega) = \frac{exp(-\sum_{c \in C} V_c(\omega))}{z} \tag{4}$$

where z is a normalizing constant to confine $P(x = \omega)$ inside of $[0, 1]$, C is the set of all *cliques* and V_c is a function whose value depends only on the nodes in clique c. There are

X1	X2	X3
X4	X5	X6
X7	X8	X9

1	0	0
0	1	0
1	0	0

Figure 4: a 3-by-3 area with elements x_1 to x_9. x_i can only be 0 or 1 in binary setup. Thus this area can have 2^9 different configuration. The right part of the figure is one configuration. MRF technique is employed to show the conditional independency given the value from its neighbors.

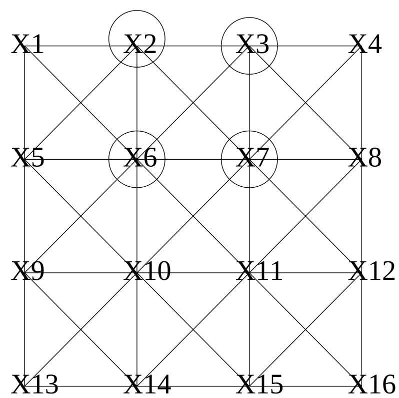

Figure 5: Illustration of *clique*. (x_2, x_7), (x_6, x_3, x_7) and (x_2, x_3, x_7, x_6) are cliques. Notice that there is no clique including 5 nodes.

broad variations of $V_c(\omega)$ playing as weighting functions for different situations. *Clique* is a set of nodes in a graph such that any pair of them are connected by an arc. We illustrate the definition by Figure 5.

2.3 Dual Lattice Number

Before get into details about Hammersley-Clifford Theorem, we introduce the concept of **Dual Lattice Number**. It is defined as virtual nodes on the edge between two adjacent pixels and only takes binary values, 0 and 1. Dual Lattice Number of 1 indicates the boundaries of groups in an image. 0 states that the two surrounding pixels belong to one group as shown in Figure 6. This procedure is called *Line Process*. There are more examples on page 735 of Sturat Geman and Donald Geman's paper.

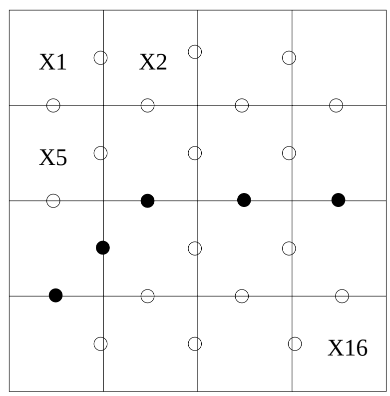

Figure 6: Line Process. Black and white disks are Dual Lattice Numbers. Black disks indicate the value of 1 and white disks represent the value of 0. After line process, above 4-by-4 area is divided into two regions.