Joint Gaussian Graphical Model Review Series – I

Probability Foundations

Beilun Wang
Advisor: Yanjun Qi

[1]Department of Computer Science, University of Virginia
http://jointggm.org/

June 23rd, 2017

Outline

Notation

Notation

$\mathbb{P}$	The probability measure.
Ω	The sample space.
$\mathcal{F}$	The event set.
X, Y, Z	The random variables.

Probability

Probability Space

Probability Space

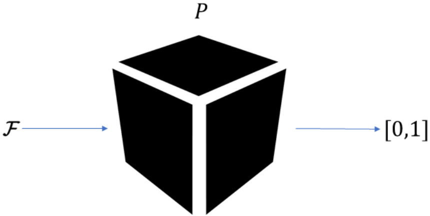
Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the probability space.

- Ω be an arbitrary non-empty set.
- $\mathcal{F} \subset 2^{\Omega}$ is a set of events.
- $\mathbb{P}$ is the probability measure. In another word, a function : $\mathcal{F} \to [0, 1]$.

Events

- $\mathcal{F}$ contains Ω.
- $\mathcal{F}$ is closed under complements.
- $\mathcal{F}$ is closed under countable unions.

Probability Measure

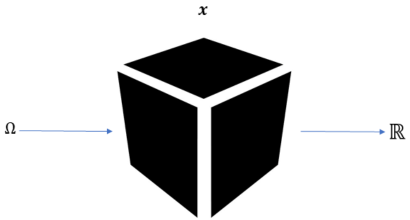


Random Variable

Random Variable

Let $X : \Omega \to \mathbb{R}$ be a random variable. X is a measurable function.

Random Variable



Probability Distribution

Probability Distribution function

Let $F(x) : \mathbb{R} \to [0, 1] = \mathbb{P}[X < x]$ where $x \in \mathbb{R}$.

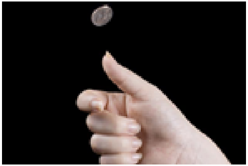
- $X = Y$, they follow same distribution?
- $F_X = F_Y$, then $X = Y$?

Joint Probability

Joint Probability

The probability distribution of random vector (X, Y).

Joint Probability



Twice

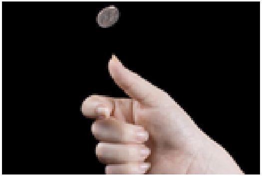
{Head, Head} {Tail, Tail} {Head, Tail}

Marginal Probability

Marginal Probability

A pair of random variable (X, Y), the probability distribution of X.

Joint Probability



Twice

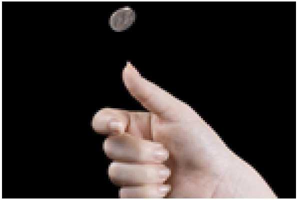
Head or Tail for the first one?

Conditional Distribution

Conditional Distribution

Given the information of Y, the probability distribution of X. Notation $X|Y$.

Joint Probability



Twice

I know the second one is Head.
Head or Tail for the first one?

Relationship

Relationship

$\mathbb{P}(X = x, Y = y) = \mathbb{P}(Y = y)\mathbb{P}(X = x | Y = y)$

Dependence and Correlation

Independence

Independence

X and Y are independent if and only if $p_{X,Y}(x, y) = p_X(x)p_Y(y)$, where p is the probability density function.

Independence

$Y|X = Y$

- Filp coin example
- Causal relationship

Correlation

Covariance

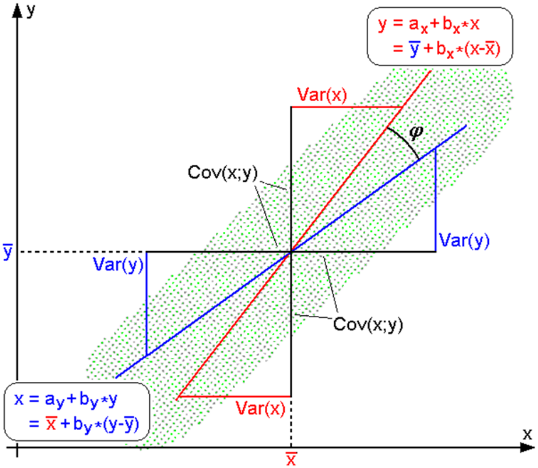
$Cov(X, Y) = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)]$, where μ_X, μ_Y is the mean vector.

Correlation

$\rho(X, Y) = \frac{Cov(X,Y)}{\sigma_X \sigma_Y}$

- Linear relationship
- Linear dependency between X and Y.
- $\rho(X, Y) = 1$ means that X and Y are in the same linear direction while $\rho(X, Y) = -1$ means that X and Y are in the reverse linear direction.
- $\rho(X, Y) = 1$ means that when X increase, Y increase with all the points lying on the same line.
- $\rho(X, Y) = 0$ means that X and Y are perpendicular with each other.

Correlation



Dependence and Correlation

- Correlation is easy to estimate the value while independence is a relationship to infer.
- Dependence is stronger relationship than correlation.
- In another word, if X and Y are independent, $\rho(X, Y) = 0$. However, the reverse doesn't hold.
- For example, suppose the random variable X is symmetrically distributed about zero and $Y = X^2$.

Gaussian Example

The distribution of bivariate Gaussian is:

$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)} * \left(\frac{(x-\mu_X)^2}{\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2} - \tag{3.1}$$

Gaussian Example

Suppose (X, Y) are uncorrelated. i.e.,$(X, Y) \sim N(0, diag(\sigma_X^2, \sigma_Y^2))$.

$$
\begin{aligned}
f(x,y) &= \frac{1}{2\pi\sigma_X\sigma_Y}\exp(-\frac{1}{2}(\frac{(x-\mu_X)^2}{\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2})) \\
&= \frac{1}{\sqrt{2\pi}\sigma_X}\exp(-\frac{1}{2}\frac{(x-\mu_X)^2}{\sigma_X^2})\frac{1}{\sqrt{2\pi}\sigma_Y}\exp(-\frac{1}{2}\frac{(y-\mu_Y)^2}{\sigma_Y^2}) \qquad (3.2) \\
&= f(x)f(y)
\end{aligned}
$$

Therefore, if (X, Y) follows bivariate Gaussian, (X, Y) are uncorrelated if and only if (X, Y) are independent.

Summary

- Correlation is easy to estimate the value while independence is a relationship to infer.
- In the Gaussian Case, they are equivalent.
- From the structure learning angle, dependence is about the causal relationship, while correlation is, more specifically, the linear relationship.

Conditional Dependence and Partial Correlation

Conditional Dependence

Let's consider a more complicated case. There is another third random variable Z. There are two ways to view the conditional dependence.

- X and Y are independent conditional on Z
- $X|Z$ and $Y|Z$ are independent

Conditional Dependence

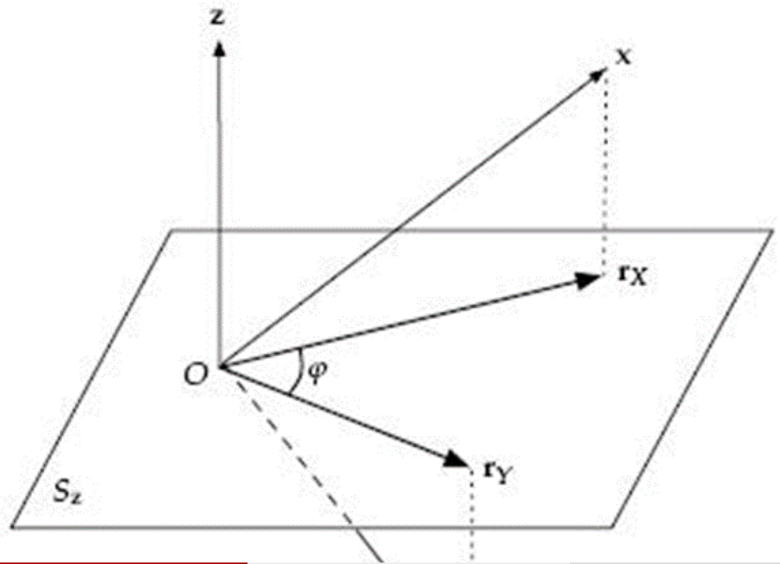
X and Y are independent on Z if and only if $p_{X,Y|Z}(x,y) = p_{X|Z}(x)p_{Y|Z}(y)$, where p is the probability density function.

Partial Correlation

Partial Correlation

Formally, the partial correlation between X and Y given random variable Z , written $\rho_{XY\cdot Z}$, is the correlation between the residuals R_X and R_Y resulting from the linear regression of X with Z and of Y with Z, respectively.

Partial Correlation



Partial Correlation Calculation

Suppose $P = \Sigma^{-1}$ (Σ is covariance matrix or Correlation matrix)

$\rho_{X_i X_j \cdot \mathbf{V} \setminus \{X_i, X_j\}} = -\frac{p_{ij}}{\sqrt{p_{ii} p_{jj}}}$.

The value is exactly related to the precision matrix (the inverse of covariance matrix)!

Conditional Dependence and Partial Correlation

- Similarly, in the Gaussian Case, they are equivalent.
- A detailed derivation is in the next talk.

Gaussian Case

- Partial Correlation is easy to estimate the value while conditional independence is a relationship to infer.
- Conditional Dependence is stronger relationship than partial correlation.
- In another word, if $X|Z$ and $Y|Z$ are independent, $\rho(X, Y \cdot Z) = 0$. However, the reverse doesn't hold.

Summary

- Partial correlation is easy to estimate the value while conditional independence is a relationship to infer.
- In the Gaussian Case, they are equivalent.
- From the structure learning angle, conditional dependence is about the causal relationship, while partial correlation is, more specifically, the linear relationship.