

# **Multivariate Probability and Related Properties**

Wei-Chang Lee, Chi-Ning Chou

November 22, 2015

# Contents

1

2

# Chapter 1

Statistical Inference I

Prof. Chin-Tsang Chiang

## Lecture Notes 17

November 22, 2015

Scribe: Wei-Chang Lee, Chi-Ning Chou

For now, we have learned how to use a single random variable to construct a model. And we have introduced several well-studied distributions and analyze their good properties. But, is that enough? Can we use only one random variable to model everything? The answer is clearly negative. But now we need to ask what are the differences between single random variable and more than one random variables? What can we benefit from that and what are the main assumptions and structures?

We start from defining random vector.

**Definition 1 (random vector)** A  $p$ -dimensional random vector is a function  $X = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix}$  from  $\Omega$  to  $\mathbb{R}^p$  such that each component is a random variable.

**Definition 2 (joint distribution)** We say  $F_X$  is a joint distribution of  $X$  such that  $F_X(\underline{x}) = P[X_1 \leq x_1, \dots, X_p \leq x_p]$ .

Now, we might wonder, what if some components are discrete and some are continuous? How to define the joint distribution?

**Example:**  $X_1$  is discrete and  $X_2$  is continuous.

- Only  $X_1$ :  $f_{X_1}(x_1) = P(X_1 = x_1)$
- Only  $X_2$ :  $f_{X_2}(x_2) = \frac{\partial}{\partial x_2} F_{X_2}(x_2)$
- Both  $X_1$  and  $X_2$ :  $\lim_{\Delta_2 \rightarrow 0} \frac{P[X_1 = x_1, X_2 \in [x_2 - \frac{\Delta_2}{2}, x_2 + \frac{\Delta_2}{2}]]}{\Delta_2}$

Finally, we give an example to show the benefit of using random vector instead of using single random variable.

**Example:** Consider the regression problem as follow. We have a dependent variable  $y$  and some explanatory variables  $X_1, \dots, X_p$ . Now we want to know the distribution of  $Y$  conditioned on  $X_1, \dots, X_p$  or the distribution of  $X_1, \dots, X_p$  conditioned on  $Y$ . To characterize these, we have to look into the correlation between  $Y$  and  $X_1, \dots, X_p$  which can be simply modeled by joint distribution.

To sum up, random vector and joint distribution provide us a structure that can describe the correlation among different sources.