

# Higher Theory of Statistics

## Math2901 UNSW

Hussain Nawaz  
hussain.nwz000@gmail.com

2022T2

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
1.1	Experiments, Sample Space and Events . . . . .	2
1.2	Sigma Algebra . . . . .	2
1.3	Conditional Probability and Independence . . . . .	3
1.4	Descriptive Statistics and R . . . . .	4
<b>2</b>	<b>Random Variables</b>	<b>4</b>
2.1	Random Variables . . . . .	4
2.2	Expectation and Variance . . . . .	6
2.3	Moment Generating Functions . . . . .	7
<b>3</b>	<b>Common Distributions</b>	<b>9</b>
3.1	Common Discrete Distributions . . . . .	9
3.2	Continuous Distributions . . . . .	10
<b>4</b>	<b>Bivariate Distributions</b>	<b>11</b>

# 1 Introduction

## 1.1 Experiments, Sample Space and Events

**Experiments** An experiment is any process that leads to a recorded observation.

**Outcome and Sample Space** An outcome is possible result of the experiment. The set of all possible outcomes is called the sample space. The sample space is often denoted by  $\Omega$ .

Observe that not all sample spaces are countable. An uncountable example would be the set of all real number between 0 and 1.

**Events** An event is a set of outcomes that is, a subset of the sample space  $\Omega$ .

**Mutual Exclusion** Events  $A, B$  are mutually exclusive (disjoint) if they have no outcomes in common. That is,  $A \cap B = \emptyset$ .

**Set Operation Revision** If you have trouble recalling the following laws, for associativity and distributivity, you may replace  $\cap$  with  $\times$  and  $\cup$  with  $+$ .

TODO: Associative and Distributive Law

## 1.2 Sigma Algebra

The  $\sigma$  algebra must be defined for rigorously working with probability. The formalization of this, is beyond the scope of this course.

The  $\sigma$ -algebra can be thought of as the family of all possible subsets or events in a sample space. Analogously, this may be conceptualised as the power-set of the sample space.

**Probability** The probability is a set function, often denoted by  $\mathcal{P}$  that maps events from the  $\sigma$ -algebra to  $[0, 1]$  and satisfies certain properties.

**Probability Space** The triplet  $\Omega, \mathcal{A}, \mathbb{P}$  is the probability space where

- $\Omega$  is the sample space,
- $\mathcal{A}$  is the  $\sigma$ -algebra,
- $\mathbb{P}$  is the probability function.

**Properties of Probability** Given the probability/sample space  $\Omega, \mathcal{A}, \mathbb{P}$ , the probability function  $\mathbb{P}$  must satisfy

- For all set  $A \in \mathcal{A}$ ,  $\mathbb{P}(A) \geq 0$
- $\mathbb{P}(\Omega) = 1$
- Countable additive. Suppose that the family of set  $A_i$

**Theorem: Continuity from below** Given an increasing sequence of events  $A_1 \subset A_2 \subset \dots \subset A_n$  then,

$$\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n \rightarrow \infty} \mathbb{P}(A_n)$$

**Theorem: Continuity from above** Given a decreasing sequence of events  $A_1 \supset A_2 \supset \dots \supset A_n$  then,

$$\mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n \rightarrow \infty} \mathbb{P}(A_n)$$

### More Probability Lemmas

- $\mathbb{P}(\emptyset) = 0$ ,
- For any  $A \in \mathcal{A}$ ,  $\mathbb{P}(A) \leq 1$  and  $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$ ,
- Suppose  $A, B \in \mathcal{A}$  and  $A \subseteq B$  then  $\mathbb{P}(A) \leq \mathbb{P}(B)$ .

## 1.3 Conditional Probability and Independence

**Conditional Probability** The conditional probability that an event  $A$  occurs given that the event  $B$  has already occurred is denoted by

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

**Independence** The events  $A$  and  $B$  are independent if  $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ .

**A lemma on independence** Given two events  $A, B$ , then

$$\mathbb{P}(A|B) = \mathbb{P}(A) \quad \text{if and only if} \quad \mathbb{P}(B|A) = \mathbb{P}(B)$$

**Pairwise Independence of Sequences** A countable sequence of events  $A_{i \in \mathbb{N}}$  is pairwise independent if

$$\mathbb{P}(A_i \cap A_j) = \mathbb{P}(A_i)\mathbb{P}(A_j) \quad \forall i \neq j.$$

**Independence of Sequences** A countable sequence of events  $A_{i \in \mathbb{N}}$  is independent if for any sub-collection  $A_{i_1}, \dots, A_{i_n}$  we have

$$\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_n}) = \prod_{j=1}^n \mathbb{P}(A_{i_j}).$$

**Multiplicative Law** Given  $A, B$  are events, then,

$$\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B)$$

This is equivalent to the multiplication down a decision tree.

**Additive Law** Let  $A, B$  be events. Then,

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B).$$

This is analogous to the inclusion-exclusion principle from set theory.

**Law of Total Probability** Suppose that  $(A_i)_{i=1, \dots, k}$  are mutually exclusive and exhaustive of  $\Omega$ . That is,

$$\bigcup_{i=1}^k A_i = \Omega.$$

Then for any event  $B$ , we have

$$\mathbb{P}(B) = \sum_{i=1}^k \mathbb{P}(B|A_i)\mathbb{P}(A_i).$$

## 1.4 Descriptive Statistics and R

**Sample Variance and Mean** Suppose that we are given observations  $x$  such that  $x = (x_1, x_2, \dots, x_n)$ .

Then, the **sample mean** is given by

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i.$$

The **sample variance** is given by

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2.$$

-

$$\frac{ds}{ds} \frac{df}{df} \frac{4}{5} \frac{4}{3} a + \frac{b}{7} \frac{a+b+(ddf)}{7} X = a_1 + a + X = a_1 + a_2 x = a_{22} - b_{22} +$$

This is the  $n$ -th This is  $x$  cringe.

.

## 2 Random Variables

### 2.1 Random Variables

**Definition: Random Variables** Suppose that we work on a probability space  $(\Omega, \mathcal{A}, \mathbb{P})$ . And the outcomes in  $\Omega$  are denoted by  $\omega$ .

Then, a random variable (r.v)  $X$  is a function from  $\Omega$  to  $\mathbb{R}$  such that  $\forall x \in \mathbb{R}$ , the set  $A_x = \{\omega \in \Omega, X(\omega) \leq x\}$ . That is, a random variable is a function that maps *Omega* to some space.

**Convention on Random Variables** Random variables are often denoted by capital letters while, the outcomes are denoted by the lower-case equivalent of the random variable.

**Cumulative Distributive** The cumulative distribution of a r.v  $X$  is defined by

$$F_X(x) = \mathbb{P}(\{\omega : X(\omega) \leq x\}) = \mathbb{P}(X \leq x).$$

**Cumulative Distribution Theorems** Suppose that  $F_X$  is cumulative distribution function of  $X$ . Then,

- It is bounded between zero and one and

$$\lim_{x \downarrow -\infty} = 0 \quad \text{and} \quad \lim_{x \uparrow \infty} = 1.$$

- It is non-decreasing. That is, if  $x \leq y$  then,  $F_X(x) \leq F_X(y)$ .
- For any  $x \leq y$ ,

$$\mathbb{P}(x < X < y) = \mathbb{P}(X \leq y) - \mathbb{P}(X \leq x) = F_X(y) - F_X(x).$$

- It is right continuous. That is,

$$\lim_{x \uparrow \infty} F_X \left( x + \frac{1}{n} \right) = F_X(x).$$

- it has finite left-hand limit and

$$\mathbb{P}(X < x) = \lim_{n \rightarrow \infty} F_x \left( x - \frac{1}{n} \right),$$

denoted by  $F_X(x-)$ . It is useful to observe that,

$$\mathbb{P}(X = x) = F_X(X) - F_X(x-) := F_X(x).$$

**Discrete Random Variables** A r.v. is said to be discrete if the image of  $X$  consists of countable many values  $x$  where  $\mathbb{P}(X = x) > 0$ . The probability function is  $\Delta F_X(x) = \mathbb{P}(X = x)$  and satisfies

$$\sum_{\text{all } x} \mathbb{P}(X = x) = 1.$$

**Continuous Random Variables and Probability Density Functions** A r.v is continuous if the image of  $X$  takes a continuum of values.

The probability density function of a r.v is a real-valued function  $f_x$  on  $\mathbb{R}$  with the property that

$$\mathbb{P}(X \in A) = \int_A f_x(y) dy,$$

for any *Borel* subset of  $\mathbb{R}$

**Required Properties of a Density Function** Valid density functions  $f : \mathbb{R} \rightarrow \mathbb{R}$  must satisfy the following properties:

- $f(x) \geq 0, \forall x \in \mathbb{R}$
- $\int_{-\infty}^{\infty} f(x)dx = 1.$

**Useful Properties of a Continuous Random Variable** For all continuous random variables  $X$ , with density  $f_x$ ,

1. If  $A = (-\infty, x]$  and creating a cumulative distribution function  $F_x$  such that  $F_X(x) = \mathbb{P}(X \in A) = \mathbb{P}(X \leq x)$  then,

$$F_X(x) = \int_{-\infty}^x f_x(y)dy.$$

2. For all  $a < b$ ,

$$\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a) = \int_a^b f_X(x)dx.$$

3. By the fundamental theorem of calculus and property 1,

$$F'_X(x) = \frac{d}{dx} \int_{-\infty}^x f_x(y)dy = f_X(x).$$

## 2.2 Expectation and Variance

**Expectation** The expectation of a r.v  $X$ , denoted by  $\mathbb{E}(X)$  may be computed depending on when  $X$  is discrete or continuous.

**Expectation of Discrete Random Variables** If  $X$  is a discrete random variable then,

$$\mathbb{E}(X) := \sum_{\text{all } x} x\mathbb{P}(X = x) = \sum_{\text{all } x} x\Delta F_x(x).$$

**Expectation of continuous Random Variables** If  $X$  is a continuous random variable then,

$$\mathbb{E}(X) := \int_{-\infty}^{\infty} xf_X(x)dx$$

**Interpreting the Expectation** Often  $\mathbb{E}(x)$  is called the *mean* of  $X$ . Observe that mean and average are not necessarily the same.  $\mathbb{E}(X)$  may be thought as the long-run average of the outcomes of  $X$ . That is, the average observation of  $X$  converges to  $\mathbb{E}(X)$ .

Where our density function represents a physical model,  $\mathbb{E}(X)$  is equivalent to the center of mass.

**Linearity of the Expectation** We note that the expectation is linear. That is, for all constants  $a, b \in \mathbb{R}$ ,

$$\mathbb{E}(aX + b) = a\mathbb{E}(X) + b.$$

**Variance** Let  $X$  be a r.v and set  $\mu = \mathbb{E}(x)$ . Then,

$$\text{Var}(X) := \mathbb{E}((X - \mu)^2).$$

The standard deviation is the square root of variance.

**Properties of Variance** Given a random variance  $X$  then, for any constants  $a, b \in \mathbb{R}$ ,

1.  $\text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$ .,
2.  $\text{Var}(aX) = a^2\text{Var}(X)$ ,
3.  $\text{Var}(X + b) = \text{Var}(X)$ ,
4.  $\text{Var}(b) = 0$ .

**Covariance** Recall that variance is  $\mathbb{E}((X - \mu)^2)$ . Suppose that  $X, Y$  are random variables such that  $\mathbb{E}(X) = \mu_X, \mathbb{E}(Y) = \mu_Y$ . Then, the covariance is defined as

$$\text{Cov}(X, Y) = \mathbb{E}((X - \mu_X)(Y - \mu_Y)).$$

Observe that  $\text{Cov}(X, X) = \text{Var}(X)$ .

**Variance and Linearity** The variance is not linear. That is,

$$\text{Var}(X + Y) \neq \text{Var}(X) + \text{Var}(Y).$$

However,

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) - 2\text{Cov}(X, Y).$$

## 2.3 Moment Generating Functions

**Moments** A moment of the random variable is denoted by

$$\mathbb{E}[X^r], \quad r = 1, 2, \dots$$

Moments measure mean, variance, skewness, and kurtosis, all ways of looking at the shape of the distribution.

Suppose that  $f(x)$  is a probability density function. Then,

$$\mathbb{E}[X^r] = \int_{-\infty}^{\infty} x^r f(x) dx$$

**Kurtosis** The kurtosis is the standard's 4th moment. It measures how *fat* the tail is. A positive kurtosis implies a thinner tail than negative kurtosis.

**Moment Generating Function** A moment generating function (MGF) is denoted as

$$M_X(u) = \mathbb{E}(e^{uX}) = \int_{\text{all } x} e^{ux} f_X(x) dx.$$

We say that the MGF of  $X$  exists if  $M_X(u)$  is finite in some interval containing zero.

**Using Moment Generating Function to Find Moments** Suppose that the moment generating function exists. Then,

$$\mathbb{E}(X^r) = \lim_{u \rightarrow 0} M_X^{(r)}(u) =: \lim_{u \rightarrow 0} \frac{d^r}{du^r} M_X(u).$$

**Equivalence of Moment Generation Functions** Let  $X, Y$  be two random variables and suppose that  $M_X(u) = M_Y(u)$  for all  $u$  in some interval containing 0. Then,

$$F_X(x) = F_Y(x), \forall x \in \mathbb{R}.$$

That is, a moment generating function (when it exists), uniquely characterises a cumulative distribution function of a random variable.

**Existence of Moments and Moment Generating Functions** If the moment generating function exists then all moments can be computed. However, the converse is not necessarily true. That is, if all the moments exist and are finite, this does not imply the moment generating function exists.

## Useful Inequalities

**Markov Inequality - Chebychev's First Inequality** For all non-negative r.v  $X$ , for  $a > 0$ ,

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(X)}{a}.$$

Equivalently,

$$\int_a^\infty f(x) dx \leq \int_{-\infty}^\infty x f(x) dx.$$

**Chebychev's Second Inequality** Suppose that  $X$  is any r.v with  $\mathbb{E}(X) = \mu$ ,  $\text{Var}(x) = \sigma^2$  and  $k > 0$ . Then

$$\mathbb{P}(|X - \mu| > \sigma) \leq \frac{1}{k^2}.$$



**Convex and Concave Functions** In probability, we may want to know if a function is concave but, cannot use the usual method of the second derivative as the function is not necessarily twice differentiable.

A function  $h$  is convex if for any  $\lambda \in [0, 1]$  and  $x_1, x_2$  in the domain of  $h$ ,

$$h(\lambda x_1 + (1 - \lambda)x_2) \leq (\geq) \lambda h(x_1) + (1 - \lambda)h(x_2).$$

**Jensen's Inequality** Let  $X$  be a random variable. Suppose that  $h$  is a convex function. Then

$$h(\mathbb{E}(X)) \leq \mathbb{E}(h(X)).$$

If  $h$  is concave then,

$$h(\mathbb{E}(X)) \geq \mathbb{E}(h(X)).$$

**Applications of Jensen's Inequality** Using Jensen's inequality, it can be shown that

$$\text{Arithmetic Mean} \geq \text{Geometric Mean} \geq \text{Harmonic Mean}.$$

## 3 Common Distributions

### 3.1 Common Discrete Distributions

**Bernoulli Distributions** A Bernoulli trial is an experiment with two outcomes; success and failure. A random variable  $X$  is defined with

$$X = \begin{cases} 1 & \text{if success,} \\ 0 & \text{if failure.} \end{cases}$$

Let  $p \in [0, 1]$  be probability of success. Then, we denote  $X \sim \text{Bernoulli}(p)$

1.  $\mathbb{P}(X = 1) = p,$
2.  $\mathbb{P}(X = 0) = 1 - p,$
3.  $\mathbb{E}(X) = p,$
4.  $\text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = 1 \times p - p^2 = p(1 - p).$

**Binomial Distribution** When there are  $n$  independent bernoulli trials with a success rate of  $p$ , and  $X :=$  total number of successes. Then,  $X$  is a Binomial r.v with parameter  $n$  and  $p$  such that we write  $X \sim \text{Bin}(n, p)$ .

Let  $(Y_i)_{i=1, \dots, n}$  be a sequence of independent bernoulli trials with success rate  $p$ . Then

$$X := \sum_{i=1}^n Y_i \text{ is } \text{Bin}(n, p).$$

Expectation exists as

$$\mathbb{E}(X) = \mathbb{E}\left(\sum_{i=1}^n Y_i\right) = \sum_{i=1}^n \mathbb{E}(Y_i) = np.$$

Alternatively, using combinatorics,

$$\mathbb{P}(X = k) = C_k^n p^k (1-p)^{n-k}.$$

**Poisson Distribution** A random variable  $X$  follows the poisson distribution with parameter  $\lambda$  if its probability function is

$$\mathbb{P}(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}.$$

Observe that

$$\mathbb{E}(X) = \lambda.$$

**Use** The poisson distribution is used to model count data. That is, counting the number of times an event occurs within a time period. The parameter  $\lambda$  represents the average number of times the event occurs in the time period of interest.

**Hypergeometric Distribution** A random variable has hypergeometric distribution with parameter  $N, m, n$  and is written as  $X \sim \text{Hyp}(n, m, N)$  if

$$\mathbb{P}(X = k) = \frac{C_x^m \times C_{n-x}^{N-m}}{C_n^N}, \quad \text{where } x = 1, \dots, n.$$

**Example of Hypergeometric** Given a box of  $N$  balls,  $m$  are red and  $N - m$  are black. Draw  $n$  balls at random and let  $X$  be the number of red balls drawn. Then,  $X \sim \text{Hyp}(n, m, N)$ .

**Remark** The “I feel like skipping this... discrete problems are not very interesting” was stated while covering this. Interpretation of this is left as an exercise to the reader.

## 3.2 Continuous Distributions

**Gaussian - Normal Random Variable** Given parameters  $\mu, \sigma^2$  has a probability density function as

$$f_X(x) = \frac{a}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}.$$

The expectation and variance are  $\mu, \sigma^2$  respectively.

**Linear Transforms** Let  $X$  be a r.v with a probability density function  $f_X$ . Let  $y = a + bX$  then for  $b > 0$  and  $a \in \mathbb{R}$ , then

$$f_Y(x) = \frac{1}{b} f_X\left(\frac{x-a}{b}\right).$$

**Linear Transformation of Normally Distributed Random Variable** Suppose that  $X \sim \mathcal{N}(\mu, \sigma^2)$  and  $a \in \mathbb{R}$  and  $b > 0$ . Then, the random variable  $Y := a + bX$  is normally distributed as

$$Y \sim \mathcal{N}(a + b\mu, b^2\sigma^2).$$

That is, normally distributed random variable are closed under linear transformation.

**Indicator Function** An indicator function of a set  $A$  is defined by

$$I_A(X) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}.$$

Commonly, we may see indicator functions over intervals that follow the notation  $I_{[a,b]}$ .

The indicator unifies expectation and probability since, the probability is the expectation of the indicator function. Therefore, it may be written that

$$\mathbb{P}(X \in A) = \int_A f_X(x)dx = \int_{-\infty}^{\infty} I_A(x)f_X(x)dx = \mathbb{E}(I_A(X)).$$

## 4 Bivariate Distributions

**Definition: Bivariate Distribution** The joint density function of two continuous random variables  $X, Y$  is given by a bivariate function  $f_{X,Y}$  with the following properties

1.  $f(x, y) \geq 0, \forall (x, y) \in \mathbb{R}^2$ .

- 2.

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y)dx dy = 1.$$

3. For any (measurable) sets  $A, B$ ,

$$\mathbb{P}(X \in A, Y \in B) = \int_A \int_B f_{X,Y}(x, y)dx dy.$$

**Min and Max Notation** We may write  $a \vee b = \max(a, b)$  and  $a \wedge b = \min(a, b)$ .

### Revision: Double Integral Theorems

1. **Tonelli's Theorem:** Suppose  $f : \mathbb{R}^2 \rightarrow \mathbb{R}_+$ . Then,

$$\int_{\mathbb{R}} \int_{\mathbb{R}} f(x, y)dx dy = \int_{\mathbb{R}} \int_{\mathbb{R}} f(x, y)dy dx.$$

2. **Fubini's Theorem:** Suppose that  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ . If

$$\int_{\mathbb{R}} \int_{\mathbb{R}} |f(x, y)| dx dy < \infty, \text{ or } \int_{\mathbb{R}} \int_{\mathbb{R}} |f(x, y)| dy dx < \infty,$$

Then,

$$\int_{\mathbb{R}} \int_{\mathbb{R}} f(x, y)dx dy < \infty, = \int_{\mathbb{R}} \int_{\mathbb{R}} f(x, y)dy dx < \infty, .$$

**Lemma: Expected Value of Bounded Borel Function** For any bounded Borel function  $g : \mathbb{R} \rightarrow \mathbb{R}$  and random variables  $X, Y$ , if the following sums / integrals are finite then,

$$\mathbb{E}(g(X, Y)) = \begin{cases} \sum_{\forall x} \sum_{\forall y} g(x, y) \mathbb{P}(X = x, Y = y) & \text{discrete ,} \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f_{X,Y}(x, y) dx dy. & \end{cases}$$

**Marginal Probability Density** This is the name of the individual density functions of a joint density function. That is,  $f_X, f_Y$  are the marginal densities of  $f_{X,Y}$ .

The marginal densities are given as

$$f_X(x) = \int_{\mathbb{R}} f_{x,y} dy..$$

Similarly, for the discrete case,

$$\mathbb{P}(X = x) = \sum_{\forall y} \mathbb{P}(X = x, Y = y)..$$

**Independence** Two random discrete variables  $X, Y$  are independent if

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

Similarly, for the continuous case, with a join probability of  $f_{X,Y}$ ,

$$f_{X,Y} = f_x(x) f_y(y), \forall x, y.$$

**Independent Product of Expectation** If  $X, Y$  are independent for bounded functions  $f, g : \mathbb{R} \rightarrow \mathbb{R}$  then

$$\mathbb{E}(g(X) f(Y)) = \mathbb{E}(g(X)) \mathbb{E}(f(Y)).$$

**Conditional Probability** Conditional probability for functions extends the standard case of sets. That is, in the discrete case,

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

Similarly for the continuous case,

$$f_{x|y} := \frac{f_{x,y}(x, y)}{f_y(y)}.$$

**Multivariate Gaussian** A random vector  $X = (X_1, X_2)$  is said to be Gaussian with  $\mu_X = (\mu_{x_1}, \mu_{x_2})$  and covariance matrix  $V$  if

$$f_X(x) = \frac{1}{\sqrt{(2\pi)^d |V|}} \exp \left( -\frac{1}{2} (X - \mu_X)^T V^{-1} (X - \mu_X) \right).$$

$d = 2$  (dimensions),  $V^{-1}$  is the matrix inverse of  $V$  and  $|V|$  is the determinant of  $V$ .

**Variance Matrix** The variance matrix is a symmetric matrix with entries

$$V_{ij} = \text{Cov}(X_i, X_j), \text{ where } i, j \in 1, \dots, d.$$

If  $X = (X_1, X_2)$  is multivariate Gaussian then  $(X_i)$  for  $i = 1, 2$  must be one-dimensional Gaussian but, the converse is not true.

**Conditional Variance and Expection** Given any bound Borel function  $g$ , the conditional expection of  $g(x)$  given the set  $\{Y = y\}$  is

$$\mathbb{E}(g(x)|Y = y) = \begin{cases} \sum_x g(x)\mathbb{P}(X = x|Y = y) & \text{discrete,} \\ \int_{-\infty}^{\infty} g(x)f_{X|Y}(x|y)dx & \text{continuous.} \end{cases}.$$

Conditional variance follows such that

$$\text{Var}(X|Y = y) = \mathbb{E}(X^2|Y = y) - (\mathbb{E}(X|Y = y))^2.$$

**Independent Conditional Variance and Expection** Let  $X, Y$  be independent random variables. Then,

$$\mathbb{E}(X|Y = y) = \mathbb{E}(X) \text{ and } \text{Var}(X|Y = y) = \text{Var}(X).$$

**Covariance** Let  $X, Y$  be random variables. Then,

$$\text{Cov}(X, Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))).$$

**Properties of Covariance** For random variables  $X, Y$ ,

1.  $\text{Cov}(X, X) = \text{Var}(X)$ .
2.  $\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$ .
3.  $\text{Cov}(X, Y) = 0$  if  $X, Y$  are independent. However a zero covariance does not imply independence.
4. The covariance is bilinear such that  $\forall a, b, c \in \mathbb{R}$ , and random variables  $X, Y, Z$ ,

$$\text{Cov}(aX + bY, Z) = a\text{Cov}(X, Z) + b\text{Cov}(Y, Z).$$

The same principle holds for  $\text{Cov}(X, aY + bZ)$ .