

# STA237 Notes

## Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
1.1	Basic Definitions . . . . .	3
1.2	Properties of Events . . . . .	3
1.2.1	Axioms . . . . .	4
1.3	Tools for Counting Sample Points . . . . .	4
<b>2</b>	<b>Conditional Probability</b>	<b>4</b>
2.0.1	Multiplication Rule . . . . .	5
2.1	Independent Events . . . . .	5
2.2	Partitions . . . . .	5
2.2.1	The Law of Total Probability . . . . .	5
2.3	Bayes' Theorem . . . . .	5
<b>3</b>	<b>Random Variables</b>	<b>6</b>
3.0.1	Result . . . . .	6
3.1	Expected Values of Random Variables . . . . .	6
3.1.1	Variance of Random Variables . . . . .	6
3.1.2	Results . . . . .	7
3.2	Distribution Function . . . . .	7
3.3	Bernoulli Distributions . . . . .	7
3.3.1	Results . . . . .	8
3.3.2	Probability Mass Functions . . . . .	8
3.4	Binomial Distributions . . . . .	8
3.4.1	Properties of Binomial Distribution . . . . .	8
3.5	Geometric Distribution . . . . .	9
3.5.1	Properties of Geometric Distribution . . . . .	9
3.5.2	Results of Geometric Probability Distribution . . . . .	9
3.6	Hypergeometric Random Variables . . . . .	9
3.6.1	Hypergeometric Probability Mass Function . . . . .	9
3.7	Poisson Probability Distribution . . . . .	9
<b>4</b>	<b>Continuous Random Variables</b>	<b>10</b>
4.1	Distribution Functions . . . . .	10
4.1.1	Properties of Distribution Functions . . . . .	10
4.2	Probability Density Function . . . . .	10
4.2.1	Properties of Density Functions . . . . .	10
4.2.2	Results . . . . .	10
4.3	Expected Values for Continuous Random Variables . . . . .	10
4.3.1	Results . . . . .	11
4.4	Variance in Continuous Random Variables . . . . .	11
4.5	Uniform Probability Distribution . . . . .	11
4.5.1	Results . . . . .	11
4.6	Normal Probability Distribution . . . . .	12

4.6.1	Results	12
4.6.2	Standard Normal Distribution	13
4.7	Normal Approximations of Binomial Distribution	13
4.8	Gamma Distribution	13
4.8.1	Results	14
4.9	Exponential Distribution	15
4.9.1	Results	15
<b>5</b>	<b>Multivariate Probability Distributions</b>	<b>16</b>
5.1	Bivariate Probability Distributions	16
5.1.1	Results	16
5.1.2	Joint Probability Density	16

# 1 Introduction

## 1.1 Basic Definitions

1. Scientific Question - A question created by an experimenter.
2. Experiment - A task to collect information in order to answer a scientific question.
3. Sample Space ( $\Omega$ ) - The set of all possible outcomes or results of an experiment.  
For example,  $\Omega = \{H, T\}$  is the sample space of tossing a coin.
4. Subsets of the sample space are called events.  
Events all use typical set operations (complements, union, intersection, etc.).

## 1.2 Properties of Events

1. We call events  $A, B$  mutually exclusive if  $A, B$  have no outcomes in common. That is,  $A \cap B = \emptyset$
2. **Demorgan's Law** - For any two events  $A, B$ , we have  $(A \cup B)^c = A^c \cap B^c$ , and  $(A \cap B)^c = A^c \cup B^c$ .
3. A **Probability Function** ( $P$ ) on a finite sample space  $\Omega$  assigns to each event in  $A$  in  $\Omega$  a number  $P(A)$  in  $[0, 1]$  such that:
  - (a)  $P(\Omega) = 1$ , and
  - (b)  $P(A \cup B) = P(A) + P(B)$ , if  $A, B$  are disjoint.  
The number  $P(A)$  is the probability for which  $A$  occurs.

Suppose we had two events  $A, B$ , and  $P(A) \cap P(B) \neq \emptyset$ . We have:

- (a) Elements of ONLY  $A$ :  $A \cap B^c$
- (b) Elements of  $A$  AND  $B$ :  $A \cap B$
- (c) Elements of ONLY  $B$ :  $B \cap A^c$

Then:

- (a)  $P(A) = P(A \cap B^c) + P(A \cap B)$
- (b)  $P(B) = P(B \cap A^c) + P(A \cap B)$
- (c)  $P(A \cup B) = P(A \cap B^c) + P(A \cap B) + P(B \cap A^c)$   
Then:  $P(A \cup B) = P(A) - P(A \cap B) + P(A \cap B) + P(B) - P(A \cap B)$   
 $= P(A) + P(B) - P(A \cap B)$

Therefore, we have  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ .

We know that  $P(A) \subseteq P(\Omega)$ , and the complement  $A^c$  is mutually exclusive.  $P(\Omega) = 1$ , and thus:

$$P(\Omega) = 1 = P(A^c) + P(A)$$

Therefore:  $P(A^c) = 1 - P(A)$ .

4.  $A$  and  $B$  are **independent** if  $P(A \cap B) = P(A) \cdot P(B)$ .

### 1.2.1 Axioms

Suppose  $\Omega$  is a sample space associated with an experiment. To every event  $A$  in  $\Omega$ , we assign a number  $P(A)$  (called the probability of  $A$ ), so that the following axioms hold:

1. Axiom 1:  $P(A) \geq 0$
2. Axiom 2:  $P(S) = 1$
3. Axiom 3: If  $A_1, A_2, \dots, A_n$  form a sequence of pairwise mutually exclusive events in  $\Omega$  (that is,  $A_i \cap A_j = \emptyset$  if  $i \neq j$ ), then

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = \sum_{i=1}^n P(A_i)$$

### 1.3 Tools for Counting Sample Points

With  $m$  elements  $a_1, a_2, \dots, a_m$ , and  $b_1, b_2, \dots, b_n$ , it is possible to form  $mn = m \times n$  pairs containing one element from each group.

An ordered arrangement of  $r$  distinct objects is called a **permutation**. The number of ways of ordering  $n$  distinct objects taken  $r$  at a time will be designated by the symbol  $P_r^n$ . That is:

$$P_r^n = n(n-1)(n-2)\dots(n-(r+1)) = \frac{n!}{(n-r)!}$$

The number of unordered subsets of size  $r$  chosen (without replacement from  $n$  available objects is:

$$\binom{n}{r} = \frac{P_r^n}{r!} = \frac{n!}{r!(n-r)!}$$

Sometimes it is denoted as  $C_r^n$ .

## 2 Conditional Probability

**Conditional probability** is the likelihood of an event occurring based on the occurrence of a previous event. That is, for two events  $R, L$ , the conditional probability of  $R$  given  $L$  is  $P(R|L)$ .

It is denoted by:

$$P(A|C) = \frac{P(A \cap C)}{P(C)},$$

provided  $P(C) > 0$ .

Note that  $P(R|L) + P(R^c|L) = 1$ :

$$\begin{aligned} P(R|L) + P(R^c|L) &= \frac{P(A \cap C)}{P(C)} + \frac{P(A^c \cap C)}{P(C)} \\ &= \frac{P(C)}{P(C)} \\ &= 1 \end{aligned}$$

Since they are mutually exclusive, the union of the intersections is  $P(C)$ .

For example, suppose we had the following events:

1.  $L$ : Born in a long month (31 days)  
 $L = \{Jan, Mar, May, Jul, Aug, Oct, Dec\}$ ;

2.  $R$ : Born in a month with letter  $r$   
 $R = \{Jan, Feb, Mar, Apr, Sep, Oct, Nov, Dec\}$

This means that the conditional probability of  $R$  given  $L$  is:

$$\begin{aligned} P(R|L) &= \frac{1/3}{7/12} \\ &= \frac{4}{7} \end{aligned}$$

### 2.0.1 Multiplication Rule

For any events  $A, C$ :

$$\begin{aligned} P(A|C) &= \frac{P(A \cap C)}{P(C)} \\ P(A \cap C) &= P(A|C) \cdot P(C) \end{aligned}$$

## 2.1 Independent Events

Events  $A, C$  are **independent** if and only if the probability of  $A$  is the same when we know that  $C$  has occurred. That is:

$$P(A|C) = P(A)$$

Then:

$$\begin{aligned} \frac{P(A \cap C)}{P(C)} &= P(A) \\ P(A \cap C) &= P(A) \cdot P(C) \end{aligned}$$

## 2.2 Partitions

For some positive integer  $k$ , let the sets  $B_1, B_2, \dots, B_k$  be such that:

1.  $\Omega = B_1 \cup B_2 \cup \dots \cup B_k$
2.  $B_i \cap B_j = \emptyset$ , for  $i \neq j$ .

Then, the collection of sets  $\{B_1, B_2, \dots, B_k\}$  is said to be a partition of  $\Omega$ .

### 2.2.1 The Law of Total Probability

Suppose that  $\{B_1, B_2, \dots, B_k\}$  is a partitions of  $\Omega$  such that  $P(B_i) > 0$  for  $i = 1, 2, \dots, k$ . Then, for any event  $A$ :

$$\begin{aligned} P(A) &= P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \dots + P(A|B_k)P(B_k) \\ &= \sum_{i=1}^k P(A|B_i)P(B_i) \end{aligned}$$

## 2.3 Bayes' Theorem

Suppose that  $\{B_1, B_2, \dots, B_k\}$  is a partition of  $\Omega$  such that  $P(B_i) > 0$ , for  $i = 1, 2, \dots, k$ . Then, for any event  $A$ :

$$P(B_j|A) = \frac{P(A|B_j)P(B_j)}{\sum_{i=1}^k P(A|B_i)P(B_i)}$$

### 3 Random Variables

**Discrete Variables** are variables whose values can be measured by counting.

For example, a course mark: 0, 1, 2, ..., 100

**Continuous Variables** are impossible to count and can never properly be counted.

For example, time or weighs: 25 years, 10, months, ...

**Categorical Variables** take on a finite number of possible values, assigning units of observation to particular groups on the basis of qualitative properties.

For some event with sample space  $\Omega$  taking multiple parameters (*e.g.*,  $\Omega = \{\sigma_1, \sigma_2\} : \sigma \in \{1, 2\}$ ), we can calculate the total outcome, i.e., the value of the function  $X : \Omega \rightarrow \mathbb{R}$ , given by:

$$X(\sigma_1, \sigma_2) = \sigma_1 + \sigma_2 \text{ for } (\sigma_1, \sigma_2) \in \Omega$$

We denote the event that the function  $S$  attains the value  $k$  by:

$$\{X = k\} = \{(\sigma_1, \sigma_2) \in \Omega : X(\sigma_1, \sigma_2) = k\}$$

We call  $X$  the **random variable**.

$X : \Omega \rightarrow \mathbb{R}$  is a **discrete random variable** if it takes on a finite number of values  $a_1, a_2, \dots, a_n$ , **or** an infinite number of values  $a_1, a_2, \dots$

The probability that  $X$  takes on the value  $x$ ,  $P(X = x)$  is the sum of probabilities of all sample points in  $\Omega$  that are assigned to the value  $x$  (i.e.,  $P(x) = P(X = x)$ ). We sometimes denote this as  $p(x)$ .

Then, the probability distribution of a discrete variable  $X$  can be represented by a formula, a table, or a graph that provides  $P(X = x)$  for all  $x$ .

#### 3.0.1 Result

For any discrete probability distribution, the following must be true:

1.  $0 \leq p(x) \leq 1$  for all  $x$
2.  $\sum_x p(x) = 1$ , where the summation is over all values of  $x$  with non-zero probability.

### 3.1 Expected Values of Random Variables

Let  $X$  be a discrete random variable with the probability function  $p(x)$ . Then, the expected value of  $X$ ,  $E(X)$ , is defined as:

$$E(X) = \sum_x xP(x),$$

where  $P(x) = P(X = x)$ . Note that  $E(x) = \mu = \sum_x xP(x)$ .

#### 3.1.1 Variance of Random Variables

If  $X$  is a random variable with **mean**  $E(X) = \mu$ , then the variance of a random variable  $X$  is the expected value of  $(X - \mu)^2$ . That is:

$$\sigma^2 = V(X) = E[(-\mu)^2]$$

The **standard deviation** of  $X$  is the positive square root of  $V(X)$ , or  $\sigma$ .

### 3.1.2 Results

1. Let  $X$  be a discrete random variable with probability function  $p(x)$ , and let  $c$  be a constant. Then,

$$\begin{aligned} E(c) &= \sum_x c \sum P(x) \\ &= c \cdot 1 \\ &= c \end{aligned}$$

Therefore,  $E(c) = c$ .

2. Note that for the variance:

(a)

$$\begin{aligned} V(c) &= E((c - \mu)^2) \\ &= E((c - c)^2) \\ &= 0 \end{aligned}$$

(b)

$$\begin{aligned} V(cX) &= c^2 V(X) \\ V(aX + b) &= a^2 V(X) \end{aligned}$$

3. Let  $X$  be a discrete random variable with probability function  $p(x)$ ,  $g(x)$  be a function of  $X$ , and let  $c$  be a constant. Then:

$$\begin{aligned} E(cx) &= cE(x) \\ &= E[ax + b] \\ &= aE(x) + b \end{aligned}$$

Therefore,  $E[cg(X)] = cE(g(X))$ .

4. Let  $X$  be a discrete random variable with probability function  $p(x)$ , and  $g_1(X), g_2(X), \dots, g_k(X)$  be  $k$  functions of  $X$ . Then:

$$E[g_1(X) + g_2(X) + \dots + g_k(X)] = E[g_1(X)] + E[g_2(X)] + \dots + E[g_k(X)]$$

## 3.2 Distribution Function

The distribution function  $F$  of a random variable  $X$  is the function  $F : \mathbb{R} \rightarrow [0, 1]$ , defined by:

$$F(a) = P(X \leq a) \text{ for } -\infty < a < \infty$$

## 3.3 Bernoulli Distributions

The Bernoulli distribution is used to model an experiment with only two possible outcomes, referred to as a ‘success’ and ‘failure’, usually encoded as 1 and 0. A Bernoulli Trial is the term used to describe these experiments.

A discrete random variable  $X$  has a Bernoulli distribution with parameter  $p$ , where  $0 \leq p \leq 1$ , if its probability mass function is given by:

$$P(X = 1) = p \text{ and } P(X = 0) = 1 - p$$

We denote this distribution by  $Ber(p)$ .

### 3.3.1 Results

1. We calculate the mean  $\mu$ :

$$\begin{aligned}\mu &= E(x) = \sum_x xP(x) \\ &= 0 \cdot (1-p) + 1 \cdot p \\ E(x) &= p\end{aligned}$$

Similarly,

$$\begin{aligned}E(x^2) &= \sum_x x^2P(x) \\ &= 0^2 \cdot (1-p) + 1^2 \cdot p \\ &= p\end{aligned}$$

2. We calculate the variance  $\sigma^2$ :

$$\begin{aligned}\sigma^2 &= V(X) = E(x^2) - \mu^2 \\ &= p - p^2 \\ V(x) &= p(1-p)\end{aligned}$$

For example: Suppose we flip a coin. Heads is a success (S), and Tail is a failure (F). We have  $P(S) = p$ , and  $P(F) = 1 - p$ . We denote  $X$  as the number of heads (i.e.,  $X = 0, 1$ ). Then,  $P(X = 0) = 1 - p$ , and  $P(X = 1) = p$ .

### 3.3.2 Probability Mass Functions

A probability mass function (pmf) is a function over the sample space of a discrete random variable  $X$  that shows  $P(X)$  is equal to a specific value. That is:

$$P(X = x) = p^x(1-p)^{1-x}, \text{ where } x = 0, 1$$

## 3.4 Binomial Distributions

A discrete random variable  $X$  has a binomial distribution with parameters  $n, p$ , where  $n = 1, 2, \dots$ , and  $0 \leq p \leq 1$ , if its probability mass function is given by:

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x} \text{ for } x = 0, 1, 2, \dots, n,$$

where  $\binom{n}{k} = \frac{n!}{(n-k)!k!}$ .

We denote this distribution by  $B(n, p)$ . We also have:

1.  $E(X) = np$
2.  $V(X) = np(1-p)$

Note that we have  $X \sim B(n, p)$

### 3.4.1 Properties of Binomial Distribution

1. The experiments consist of a fixed number,  $n$ , identical trials.
2. Each trial results in one of two outcomes ( $S, F$ ).
3.  $P(S) = p$  for every trial, and  $P(F) = 1 - p$ .
4. The trials are independent.



### 3.5 Geometric Distribution

A random variable  $Y$  is said to have a **geometric probability distribution** if and only if

$$p(y) = q^{y-1} \cdot p, \text{ where } y = 1, 2, 3, \dots; 0 \leq p \leq 1$$

That is,  $p(Y) = (1 - p)^{y-1} \cdot p$ .

This variable  $Y$  is the number of trials for which the first success occurs.

#### 3.5.1 Properties of Geometric Distribution

1. The random variable with the geometric probability distribution is associated with an experiment that shares some of the characteristics of a binomial experiment.
2. Each trial has two outcomes,  $S, F$ .
3.  $P(S) = p, P(F) = 1 - p$ .
4. The trials are independent.
5. We are interested in the random variable  $Y$ , which is the number of trials on which the first success occurs.

#### 3.5.2 Results of Geometric Probability Distribution

If  $Y$  is a random variable with a geometric distribution:

$$\mu = E(Y) = \frac{1}{p} \text{ and } \sigma^2 = V(Y) = \frac{1-p}{p^2}$$

### 3.6 Hypergeometric Random Variables

The hypergeometric probability distribution is a realistic model for some types of countable data. It has the following characteristics:

1. The experiment consists of randomly drawing  $n$  elements without replacement from a set of  $N$  elements;  $r$  of which are  $S$ 's, and  $N - r$  are  $F$ 's.
2. The hypergeometric random variable  $X$  is the number of  $S$ 's in the draw of  $n$  elements.

Note that both the hypergeometric and binomial characteristics stipulate that each draw or trial results in one of two outcomes. The basic differences between these random variables is that **hypergeometric trials are dependent**, while binomial trials are independent.

#### 3.6.1 Hypergeometric Probability Mass Function

We calculate the pmf of hypergeometric distributions as:

$$P(x) = \frac{\binom{r}{x} \cdot \binom{N-r}{n-x}}{\binom{N}{n}} : x = \max[0, n - (N - r)], \dots, \min[r, n],$$

where  $N$  is the total number of elements,  $r$  is the number of  $S$  in  $N$ ,  $n$  is the number of elements drawn,  $x$  is the number of  $S$  in  $n$ .

### 3.7 Poisson Probability Distribution

For a random variable  $X$ , it is said to have a Poisson probability distribution if and only if:

$$p(x) = \frac{\lambda^x e^{-\lambda}}{x!} \text{ for } x = 0, 1, 2, \dots, \lambda > 0$$

We have  $E(X) = \lambda$  and  $V(X) = \lambda$ .

## 4 Continuous Random Variables

A random variable that can take on any value in an interval is called **continuous**, and we can study probability distribution for continuous random variables.

### 4.1 Distribution Functions

Let  $Y$  denote any random variable. The **distribution function** of  $Y$ , denoted  $F(y)$ , is such that  $F(y) = P(Y \leq y)$  for  $-\infty < y < \infty$ .

A random variable  $Y$  with distribution function  $F(y)$  is **continuous** if  $F(y)$  is continuous, for  $-\infty < y < \infty$ .

#### 4.1.1 Properties of Distribution Functions

If  $F(y)$  is a distribution function, then:

1.  $F(-\infty) = \lim_{y \rightarrow -\infty} F(y) = 0$
2.  $F(\infty) = \lim_{y \rightarrow \infty} F(y) = 1$
3.  $F(y)$  is a non-decreasing function of  $y$ .  
If  $y_1, y_2$  are any values such that  $y_1 < y_2$ , then  $F(y_1) \leq F(y_2)$ .

### 4.2 Probability Density Function

Let  $F(y)$  be the distribution function for a continuous random variable  $Y$ . Then,  $f(y)$ , given by:

$$f(y) = \frac{dF(y)}{dy} = F'(y)$$

wherever the derivative exists, is called the **probability density function** for the random variable  $Y$ .

#### 4.2.1 Properties of Density Functions

If  $f(y)$  is a density function for a continuous random variable, then:

1.  $f(y) \geq 0$  for all  $y$ ,  $-\infty < y < \infty$ .
2.  $\int_{-\infty}^{\infty} f(y)dy = 1$ .

#### 4.2.2 Results

If the random variable  $Y$  has a density function  $f(y)$ , and for  $a < b$ , the probability that  $Y$  falls into the interval  $[a, b]$  is:

$$P(a \leq y) = \int_a^b f(y)dy$$

### 4.3 Expected Values for Continuous Random Variables

The expected value for a continuous random variable  $Y$  is:

$$E(Y) = \int_{-\infty}^{\infty} yf(y)dy$$

provided that the integral exists.

### 4.3.1 Results

Let  $g(Y)$  be a function of  $Y$ . Then, the expected value of  $g(Y)$  is given by:

$$\mu = E[g(y)] = \int_{-\infty}^{\infty} g(y)f(y)dy,$$

provided that the integral exists.

Additionally, let  $c$  be a constant and let  $g(Y), g_1(Y), g_2(Y), \dots, g_k(Y)$  be functions of a continuous random variable  $Y$ . Then the following results hold:

1.  $E(c) = c$
2.  $E(c \cdot g(Y)) = cE(g(Y))$
3.  $E(g_1(Y) + \dots + g_k(Y)) = E[g_1(Y)] + \dots + E[g_k(Y)]$

### 4.4 Variance in Continuous Random Variables

The variance of a random variable  $X$  is defined by:

$$\begin{aligned}\sigma &= V(X) \\ &= E(x - \mu)^2 \\ &= \int_{-\infty}^{\infty} (x - \mu)^2 f(x)dx\end{aligned}$$

This process takes some time, so we can alternatively calculate this as:

$$V(X) = E(X)^2 - \mu^2$$

Knowing this, we then have  $E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)dx$ .

### 4.5 Uniform Probability Distribution

If  $a < b$ , a random variable  $Y$  is said to have a continuous **uniform probability distribution** on the interval  $(a, b)$  if and only if the density function of  $Y$  is:

$$f(y) = \begin{cases} \frac{1}{(b-a)} & a \leq y \leq b \\ 0 & \text{elsewhere} \end{cases}$$

#### 4.5.1 Results

If  $a < b$ , and  $Y$  is a random variable uniformly distributed on the interval  $(a, b)$ , then:

1. The mean:

$$\begin{aligned}\mu &= E(Y) = \int_{-\infty}^{\infty} yf(y)dy \\ &= \int_a^b y \cdot \frac{1}{(b-a)} dy \\ &= \frac{1}{b-a} \left[ \frac{y^2}{2} \right]_a^b \\ &= \frac{b^2 - a^2}{2(b-a)} \\ &= \frac{a+b}{2}\end{aligned}$$

2. The variance:

$$\begin{aligned}
 \mu^2 = E(Y^2) &= \int_a^b y^2 \cdot \frac{1}{b-a} dy \\
 &= \frac{1}{b-a} \left[ \frac{y^3}{3} \right]_a^b \\
 &= \frac{b^3 - a^3}{3(b-a)} \\
 &= \frac{(b-a)(b^2 + ab + a^2)}{3(b-a)} \\
 &= \frac{a^2 + ab + b^2}{3}
 \end{aligned}$$

Then:

$$\begin{aligned}
 \sigma^2 = V(Y) &= E(Y^2) - \mu^2 \\
 &= \frac{a^2 + ab + b^2}{3} - \frac{a^2 + 2ab + b^2}{4} \\
 &= \frac{4a^2 + 4ab + 4b^2 - 3a^2 - 6ab + 3b^2}{12} \\
 &= \frac{a^2 - 2ab + b^2}{12} \\
 &= \frac{(b-a)^2}{12}
 \end{aligned}$$

## 4.6 Normal Probability Distribution

A random variable  $Y$  is said to have a **normal probability distribution** if and only if, for  $\sigma > 0$  and  $-\infty < \mu < \infty$ , the density function of  $Y$  is:

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\left(\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2\right)}, \quad -\infty < y < \infty$$

Then,  $Y \sim N(\mu, \sigma)$ .

### 4.6.1 Results

If  $Y$  is a normally distributed random variable with parameters  $\mu$  and  $\sigma$ , then:

1. The mean:

$$E(Y) = \mu$$

2. The variance:

$$V(Y) = \sigma^2$$

However, calculating the integrals of these are extremely different to calculate, so we can standardize normal distributions in order to approximate them.

#### 4.6.2 Standard Normal Distribution

For  $Y \sim N(\mu, \sigma)$ , we want to find the standard normal distribution  $Z$ :

$$Z = \frac{Y - \mu}{\sigma} \sim N(E(Z), V(Z))$$

We calculated the mean and variance:

$$\begin{aligned} E(Z) &= E\left(\frac{Y - \mu}{\sigma}\right) \\ &= \frac{1}{\sigma} E(Y - \mu) \\ &= \frac{1}{\sigma} (E(Y) - \mu) \\ &= \frac{\mu - \mu}{\sigma} \\ &= 0, \end{aligned}$$

and also:

$$\begin{aligned} V(Z) &= V\left(\frac{Y - \mu}{\sigma}\right) \\ &= \frac{1}{\sigma^2} V(Y - \mu) \\ &= \frac{V(Y)}{\sigma^2} \\ &= \frac{\sigma^2}{\sigma^2} \\ &= 1 \end{aligned}$$

Therefore,

$$Z = \frac{Y - \mu}{\sigma} \sim N(0, 1)$$

#### 4.7 Normal Approximations of Binomial Distribution

Recall the formula for binomial distributions:

$$P(X = x) = \binom{1000}{x} 0.02^x \cdot 0.98^{n-x}$$

If  $X = (n, p)$ , and  $n$  is large, then the distribution of  $X$  is approximately normal with mean  $np$  and variance  $np(1 - p)$ . Equivalently, the standardized random variable

$$\frac{X - \mu}{\sigma} \sim N(0, 1) \implies \frac{X - np}{\sqrt{np(1 - p)}}$$

has an approximate standard normal distribution.

#### 4.8 Gamma Distribution

A random variable  $Y$  is said to have a **Gamma distribution** with parameters  $\alpha > 0$  and  $\beta > 0$  if and only if the density of  $Y$  is:

$$f(y) = \begin{cases} \frac{y^{\alpha-1} e^{-y/\beta}}{\beta^\alpha \Gamma(\alpha)} & y \geq 0 \\ 0 & \text{otherwise.} \end{cases}$$

where:

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy$$

Note that the quantity  $\Gamma(\alpha)$  is known as the Gamma function.

Using direct integration:  $\Gamma(1) = 1$ .

Using integration by parts:  $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$ , for any  $\alpha > 1$  and  $\Gamma(n) = (n - 1)!$  for  $n \in \mathbb{Z}$ .

#### 4.8.1 Results

1. We calculate the mean  $\mu$ :

$$\begin{aligned}\mu = E(Y) &= \int_0^\infty \frac{y \cdot y^{\alpha-1} e^{-y/\beta}}{\beta^\alpha \Gamma(\alpha)} dy \\ &= \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^\infty y^\alpha e^{-y/\beta} dy\end{aligned}$$

Let  $z = \frac{y}{\beta}$ . If  $y = 0$ , then  $z = 0$ . If  $y \rightarrow \infty$ , then  $z \rightarrow \infty$ .

Then, we have  $y = \beta \cdot z \implies dy = \beta dz$ .

Hence:

$$\begin{aligned}E(Y) &= \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^\infty (\beta z)^\alpha e^{-z} \beta dz \\ &= \frac{\beta}{\Gamma(\alpha)} \int_0^\infty z^{(\alpha+1)-1} e^{-z} dz \\ &= \frac{\beta}{\Gamma(\alpha)} \Gamma(\alpha + 1) \\ &= \frac{\beta \alpha}{\Gamma(\alpha)} \cdot \Gamma(\alpha) \\ \mu &= \alpha \beta\end{aligned}$$

2. We calculate the variance  $\sigma^2$ :

$$\begin{aligned}E(Y^2) &= \int_0^\infty \frac{y^2 y^{\alpha-1} e^{-y/\beta}}{\beta^\alpha \Gamma(\alpha)} dy \\ &= \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^\infty y^{\alpha+1} e^{-y/\beta} dy \\ &= \frac{1}{\beta^\alpha \Gamma(\alpha)} \int_0^\infty (\beta z)^{\alpha+1} e^{-z} \beta dz \\ &= \frac{\beta^2}{\Gamma(\alpha)} \int_0^\infty z^{(\alpha+2)-1} e^{-z} dz \\ &= \frac{\beta^2}{\Gamma(\alpha)} \cdot \Gamma(\alpha + 2) \\ &= \frac{\beta^2 (\alpha + 1) \Gamma(\alpha + 1)}{\Gamma(\alpha)} \\ &= \frac{\beta^2 (\alpha + 1) \alpha \Gamma(\alpha)}{\Gamma(\alpha)} \\ &= \alpha^2 \beta^2 + \alpha \beta^2 \\ \sigma^2 = V(Y) &= E(Y^2) - \mu^2 \\ &= \alpha^2 \beta^2 + \alpha \beta^2 - \alpha^2 \beta^2 \\ &= \alpha \beta^2\end{aligned}$$

## 4.9 Exponential Distribution

A random variable  $Y$  is said to have an **exponential distribution** with parameter:

$$\beta > 0 \text{ if and only if the density of } Y \text{ is } f(y) = \begin{cases} \frac{1}{\beta} e^{-y/\beta} & y \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

We can then calculate the probabilities using:

$$\begin{aligned} F(y) = P(Y \leq y) &= \int_0^y \frac{1}{\beta} e^{-t/\beta} dt \\ &= \left[ e^{-t/\beta} \right]_0^y \\ &= 1 - e^{-y/\beta} \\ P(Y > y) &= e^{-y/\beta} \end{aligned}$$

We also have:

$$f(y) = \lambda e^{-\lambda y}$$

### 4.9.1 Results

1. We calculate the mean  $\mu$ :

$$\begin{aligned} \sigma = E(Y) &= \int_0^\infty y \frac{1}{\beta} e^{-y/\beta} dy \\ &= \frac{1}{\beta} \int_0^\infty y \cdot e^{-y/\beta} dy \\ &= \frac{1}{\beta} \left( \left[ \frac{ye^{-y/\beta}}{1/\beta} \right]_0^\infty - \int_0^\infty \frac{e^{-y/\beta}}{1/\beta} dy \right) \\ &= \left( \left[ ye^{-y/\beta} \right]_0^\infty - \int_0^\infty e^{-y/\beta} dy \right) \\ &= 0 - 0 + \left[ \frac{e^{y/\beta}}{1/\beta} \right]_0^\infty \\ &= \beta[1 - 0] \\ &= \beta \end{aligned} \quad \text{Using l'Hopital's}$$

2. We calculate the variance  $\sigma^2$ :

$$\begin{aligned} E(Y^2) &= \int_0^\infty y^2 \frac{1}{\beta} e^{-y/\beta} dy \\ &= \frac{1}{\beta} \left( \left[ \frac{y^2 e^{-y/\beta}}{1/\beta} \right]_0^\infty + \int_0^\infty \frac{e^{-y/\beta}}{1/\beta} 2y dy \right) \\ &= \left( \left[ y^2 e^{-y/\beta} \right]_0^\infty + \int_0^\infty e^{-y/\beta} 2y dy \right) \\ &= 0 + 2\beta \int_0^\infty \frac{ye^{-y/\beta}}{\beta} dy \\ &= 2\beta E(Y) \\ &= 2\beta^2 \\ \sigma^2 = V(Y) &= E(Y^2) - \mu^2 \\ &= 2\beta^2 - \beta^2 \\ &= \beta^2 \end{aligned}$$

## 5 Multivariate Probability Distributions

So far, we have only dealt with one-dimensional distributions of experiments and events. However, we can also consider the **weight** as well as the height of the events when conducting experiments.

### 5.1 Bivariate Probability Distributions

Let  $Y_1, Y_2$  be discrete random variables. the **joint or bivariate probability function** for  $Y_1, Y_2$ , is given by:

$$p(y_1, y_2) = P(Y_1 = y_1, Y_2 = y_2), -\infty < y_1 < \infty, -\infty < y_2 < \infty$$

We also define the **joint distribution function**:

$$F(y_1, y_2) = P(Y_1 \leq y_1, Y_2 \leq y_2), -\infty < y_1 < \infty, -\infty < y_2 < \infty$$

#### 5.1.1 Results

Let  $Y_1, Y_2$  be discrete random variables with joint probability function  $p(y_1, y_2)$ . Then:

1.  $p(y_1, y_2) \geq 0$  for all  $y_1, y_2$ .
2.  $\sum_{y_1, y_2} p(y_1, y_2) = 1$ , where the sum is over all values  $(y_1, y_2)$  that are assigned for non-zero probabilities.

#### 5.1.2 Joint Probability Density

Let  $Y_1, Y_2$  be continuous random variables with joint distribution function  $F(y_1, y_2)$ . If there is a non-negative function  $f(y_1, y_2)$  such that:

$$F(y_1, y_2) = \int_{-\infty}^{y_1} \int_{-\infty}^{y_2} f(t_1, t_2) dt_2 dt_1,$$

for all  $-\infty < y_1 < \infty$ , and  $-\infty < y_2 < \infty$ ; then  $Y_1, Y_2$  are said to be **joint continuous random variables**. The function  $f(y_1, y_2)$  is called the **joint probability density function**.

In other words, we calculate the **volume** of the function in order to calculate the probability density. It is often easier to draw a cubic surface of the probability function in order to calculate the probability density.