

# Probability and Statistics

Based on lectures by Dr. Angshuman Bhattacharya  
Notes taken by Rwik Dutta

---

These notes are not endorsed by the lecturers, and I have modified them (often significantly) after lectures. They are nowhere near accurate representations of what was actually lectured, and in particular, all errors are almost surely mine.<sup>1</sup>

## Contents

<b>1</b>	<b>Probability Space</b>	<b>2</b>
1.1	Event space . . . . .	2
1.2	Probability measure . . . . .	2
<b>2</b>	<b>Uniform Probability Spaces</b>	<b>2</b>
<b>3</b>	<b>Conditional Probability</b>	<b>3</b>
3.1	Baye's Rule . . . . .	3
3.2	Independence . . . . .	3
<b>4</b>	<b>Combinatorial Analysis</b>	<b>3</b>
4.1	Permutations . . . . .	3
4.2	Combinations . . . . .	3
<b>5</b>	<b>Discrete Random Variable</b>	<b>3</b>
5.1	Discrete Density . . . . .	4
5.2	Joint Density . . . . .	4
5.3	Independence . . . . .	4
5.4	Sum . . . . .	4
5.5	Conditional Density . . . . .	4
5.6	Probability Generating Function . . . . .	4
5.7	Expectation . . . . .	4
5.8	Variance . . . . .	5
5.9	Covariance and Correlation . . . . .	5
5.10	Schwartz and Chebyshev's Inequalities . . . . .	5
<b>6</b>	<b>Important Discrete Densities</b>	<b>6</b>
6.1	Uniform . . . . .	6
6.2	Bernoulli . . . . .	6
6.3	Binomial . . . . .	6
6.4	Geometric . . . . .	6
6.5	Negative binomial . . . . .	6
6.6	Hypergeometric . . . . .	6
6.7	Poisson . . . . .	7
6.8	Multinomial . . . . .	7
<b>7</b>	<b>Continuous Random Variable</b>	<b>7</b>
7.1	Continuous Distribution . . . . .	7
7.2	Continuous Density . . . . .	7
7.3	Change of Variable . . . . .	8
7.4	Joint distribution . . . . .	8
7.5	Joint Density . . . . .	8
7.6	Independence . . . . .	8
7.7	Sum . . . . .	8
7.8	Conditional Density . . . . .	8
7.9	Expectation . . . . .	9
7.10	Moments . . . . .	9

---

<sup>1</sup>This is how Dexter Chua describes his lecture notes from Cambridge. I could not have described mine in any better way.

<b>8</b>	<b>Important Continuous Densities</b>	<b>9</b>
8.1	Symmetric . . . . .	9
8.2	Standard Normal . . . . .	9
8.3	Normal . . . . .	9
8.4	Exponential . . . . .	9
8.5	Gamma . . . . .	10
8.6	Cauchy . . . . .	10
8.7	Standard Bivariate Normal . . . . .	10
<b>9</b>	<b>Central Limit Theorem</b>	<b>10</b>

# 1 Probability Space

## 1.1 Event space

**Definition 1** ( $\sigma$ -field). Let  $\mathcal{A}$  be a collection of subsets of  $\Omega$  ( $A \in \mathcal{A} \Rightarrow A \subset \Omega$ ).  $\mathcal{A}$  is called a  $\sigma$ -algebra (or,  $\sigma$ -field) of subsets of  $\Omega$  if

1.  $\mathcal{A}$  is non-empty.
2.  $A \in \mathcal{A} \Rightarrow A^C \in \mathcal{A}$
3. If  $A, B \in \mathcal{A}$ ,  $A \cup B \in \mathcal{A}$  and  $A \cap B \in \mathcal{A}$ .

**Theorem 1.** A  $\sigma$ -algebra  $\mathcal{A}$  of subsets of  $\Omega$  contains  $\Omega$  and  $\emptyset$ .

**Theorem 2.** The power set of  $\Omega$  ( $\Omega \neq \emptyset$ ) forms a  $\sigma$ -algebra of subsets of  $\Omega$ .

This will be assumed to be the event space, unless mentioned otherwise.

**Definition 2** (Event space, sample space). An event space  $\mathcal{A}$  of  $\Omega$  is a  $\sigma$ -algebra of subsets of  $\Omega$ . The elements of  $\mathcal{A}$  are called events.  $\Omega$  is called the sample space.

## 1.2 Probability measure

**Definition 3.** Let  $\mathcal{A}$  be an event space of  $\Omega$ .  $P : \mathcal{A} \rightarrow \mathbb{R}$  is called a probability measure of  $\mathcal{A}$  if

1.  $P(A) \geq 0, \forall A \in \mathcal{A}$
2.  $P(\Omega) = 1$
3. Let  $A, B \in \mathcal{A}$ . If  $A \cap B = \emptyset$ ,

$$P(A \cup B) = P(A) + P(B)$$

**Theorem 3.**  $P(A) \leq 1, \forall A \in \mathcal{A}$ . Hence, range of  $P$  is  $[0, 1]$ .

**Theorem 4.**  $P(A^C) = 1 - P(A)$

**Corollary 4.1.** 
$$P\left(\underbrace{\bigcup_{i=1}^{\infty} A_i}_{\text{at least one of the events occurs}}\right) = 1 - P\left(\underbrace{\bigcap_{i=1}^{\infty} A_i^C}_{\text{none of the events occur}}\right)$$

**Theorem 5.**  $A \subset B \Rightarrow P(A) \leq P(B)$

**Theorem 6.**  $P(A \cup B) \leq P(A) + P(B)$

**Theorem 7.**

1. If  $A_1 \subset A_2 \subset A_3 \subset \dots$  and  $A = \bigcup_{n \geq 1} A_n$ ,

$$P(A) = \lim_{n \rightarrow \infty} P(A_n)$$

2. If  $A_1 \supset A_2 \supset A_3 \supset \dots$  and  $A = \bigcap_{n \geq 1} A_n$ ,

$$P(A) = \lim_{n \rightarrow \infty} P(A_n)$$

**Definition 4** (Probability space). If  $\Omega$  is a sample space,  $\mathcal{A}$  is an event space of  $\Omega$  and  $P$  is a probability measure of  $\mathcal{A}$ ,  $(\Omega, \mathcal{A}, P)$  is called a probability space.

## 2 Uniform Probability Spaces

**Definition 5** (Symmetric probability space).  $(\Omega, \mathcal{A}, P)$  (where  $\Omega$  is a finite set) is called a symmetric probability space if  $P(A) = P(B)$  for all singleton sets  $A, B \in \mathcal{A}$ .

**Theorem 8.**  $P(\{x\}) = \frac{1}{|\Omega|}, \forall x \in \Omega$ . Also,  $P(A) = \frac{|A|}{|\Omega|}, \forall A \subset \Omega$ .

**Definition 6** (Cardinality of uncountable set). If  $A \subset \mathbb{R}^n$ ,  $|A|$  denotes the  $n$ -dimensional volume of the region  $A$ .

**Definition 7** (Uniform probability space). Let  $\Omega \subset \mathbb{R}^n$  where  $|\Omega|$  is finite.  $(\Omega, \mathcal{P}(\Omega), P)$  is called an uniform probability space if  $P(A) = \frac{|A|}{|\Omega|}, \forall A \subset \Omega$ .

### 3 Conditional Probability

**Definition 8** (Conditional probability). Let  $A, B$  be two events such that  $P(A) \neq 0$ . The conditional probability of  $B$  given  $A$  denoted by  $P(B|A)$  is

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

**Theorem 9** (Principle of total probability). Let  $A_k, 1 \leq k \leq n$  be mutually disjoint events in  $\Omega$  such that  $\bigcup_{k=1}^n A_k = \Omega$  and  $P(A_k) \neq 0, \forall 1 \leq k \leq n$ .

$$P(B) = \sum_{k=1}^n P(A_k)P(B|A_k), \forall B \in \mathcal{A}$$

#### 3.1 Baye's Rule

**Theorem 10** (Baye's rule). Let  $A_k, 1 \leq k \leq n$  be mutually disjoint events in  $\Omega$  such that  $\bigcup_{k=1}^n A_k = \Omega$  and  $P(A_k) \neq 0, \forall 1 \leq k \leq n$ . If  $B \in \mathcal{A}$  such that  $P(B) \neq 0$ ,

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{\sum_{k=1}^n P(A_k)P(B|A_k)}$$

#### 3.2 Independence

**Definition 9.** Two events  $A$  and  $B$  are independent iff  $P(A \cap B) = P(A)P(B)$ .

**Theorem 11.** If  $P(A) > 0$ ,  $A, B$  are independent iff  $P(B|A) = P(B)$ .

### 4 Combinatorial Analysis

**Definition 10.** (Ordered sample) A random sample of size  $r$  from population  $S$  is the  $r$ -tuple  $(x_1, x_2, \dots, x_r)$ ,  $x_i \in S$ .

It is called a random sample **without replacement** if  $x_i = x_j \Rightarrow i = j (r \leq |S|)$ . Otherwise, it is a random sampling **with replacement**.

**Theorem 12.** The number of possible random samples with replacement of size  $r$  from a population of size  $n$  is  $n^r$ .

It is the number of ways in which  $r$  balls can be picked from  $n$  distinct balls, with replacement, where the order of picking matters.

#### 4.1 Permutations

**Theorem 13** (Permutation). The number of possible random samples without replacement of size  $r$  from a population of size  $n$  is

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

It is the number of ways in which  $r$  balls can be picked from  $n$  distinct balls, where the order of picking matters.

#### 4.2 Combinations

**Theorem 14** (Combinations). The number of ways in which  $r$  balls can be picked from  $n$  distinct balls, where the order of picking does not matter is

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

**Theorem 15.** The number of  $r$ -tuples  $(x_1, x_2, \dots, x_r)$ ,  $x_i \in \mathbb{N} \cup \{0\}$  that satisfy  $(n \in \mathbb{N})$

$$x_1 + x_2 + \dots + x_r = n$$

is  $\binom{n+r-1}{r-1}$ . If  $x_i \in \mathbb{N}$ , we have  $\binom{n-1}{r-1}$  solutions.

### 5 Discrete Random Variable

**Definition 11** (Discrete random variable). A discrete real-valued random variable  $X$  on a probability space  $(\Omega, \mathcal{A}, P)$  is a function  $X$  with domain  $\Omega$  and range a finite or countably infinite subset  $\{x_1, x_2, \dots\}$  of  $\mathbb{R}$  such that  $\{\omega : X(\omega) = x_i\}$  is an event for all  $i$ .

## 5.1 Discrete Density

**Definition 12** (Discrete density).  $f_X : \mathbb{R} \rightarrow [0, 1]$ ,

$$f_X(x) = P(X = x)$$

is the discrete density of random variable  $X$ .

## 5.2 Joint Density

**Definition 13** (Discrete random vector). A discrete real-valued random vector  $\mathbf{X}$  on a probability space  $(\Omega, \mathcal{A}, P)$  is a function  $\mathbf{X}$  with domain  $\Omega$  and range a finite or countably infinite subset  $\{\mathbf{x}_1, \mathbf{x}_2, \dots\}$  of  $\mathbb{R}^r$  such that  $\{\omega : \mathbf{X}(\omega) = \mathbf{x}_i\}$  is an event for all  $i$ .

The density is the function  $f(\mathbf{x}) = P(\mathbf{X} = \mathbf{x})$  or

$$f_{X_1, X_2, \dots, X_r}(x_1, x_2, \dots, x_r) = P(X_1 = x_1, X_2 = x_2, \dots, X_r = x_r)$$

where  $X_i$  are components of  $\mathbf{X}$ .  $f_{X_1, X_2, \dots, X_r}$  is called the **joint density** of the random variables  $X_1, X_2, \dots, X_r$ .

## 5.3 Independence

**Definition 14** (Independent random variables). Two random variables  $X, Y$  are independent iff

$$f_{X,Y}(x, y) = f_X(x)f_Y(y), \forall x, y \in \mathbb{R}$$

## 5.4 Sum

**Theorem 16** (Sum of random variables).

$$f_{X+Y}(z) = P(X = x, Y = z - x) = \sum_x f_{X,Y}(x, z - x)$$

## 5.5 Conditional Density

**Definition 15** (Conditional density).  $P(Y = y|X = x) = \frac{P(X=x, Y=y)}{P(X=x)} = \frac{f_{X,Y}(x, y)}{f_X(x)}$

$$f_{Y|X}(y|x) := \begin{cases} \frac{f_{X,Y}(x, y)}{f_X(x)}, & f_X(x) \neq 0 \\ 0, & f_X(x) = 0 \end{cases}$$

is called the conditional density of  $Y$  given  $X$ .

## 5.6 Probability Generating Function

**Definition 16** (Probability generating function).  $\Phi_X : [-1, 1] \rightarrow \mathbb{R}$ ,

$$\Phi_X(t) = \sum_{x=0}^{\infty} f_X(x)t^x$$

## 5.7 Expectation

**Definition 17** (Expectation of discrete random variable). Let the range of  $X$  be  $\{x_1, x_2, \dots\}$ . The expectation of  $X$  is

$$E(X) = \sum_{i=1}^{\infty} x_i f_X(x_i)$$

**Theorem 17.** Let  $\varphi$  be defined on the range of  $X$  and  $Z = \varphi(X)$ .

$$E(Z) = \sum_i \varphi(x_i) f_X(x_i)$$

**Definition 18** (Conditional expectation). The conditional expectation of  $Y$ , given  $X$  is

$$E(Y|X = x) = \sum_i y_i f(y|x)$$

**Theorem 18** (Linearity of expectation).

$$E(cX) = cE(X)$$

$$E(X + Y) = E(X) + E(Y)$$

**Theorem 19.**  $P(X \geq Y) = 1 \implies E(X) \geq E(Y)$ . Moreover,  $E(X) = E(Y) \iff P(X = Y) = 1$ .

**Theorem 20.**  $|E(X)| \leq E(|X|)$

**Lemma 20.1.**  $P(|X| \leq M) = 1 \implies |E(X)| \leq M$

**Lemma 20.2.**  $P(|X - Y| \leq M) = 1 \implies |E(X) - E(Y)| \leq M$

**Theorem 21.** If  $X, Y$  are independent,  $E(XY) = E(X)E(Y)$

## 5.8 Variance

**Definition 19** (Moment). Let  $X$  be a random variable with expectation  $\mu$ . The  $r^{th}$  moment of  $X$  is  $E(X^r)$  and the  $r^{th}$  central moment is  $E((X - \mu)^r)$

**Definition 20** (Variance). The  $2^{nd}$  central moment.

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

$V(X)$  is denoted by  $\sigma^2$ , where  $\sigma \geq 0$  is called the **standard deviation** of the  $X$ .

**Theorem 22.**

$$V(X + b) = V(X)$$

$$V(aX) = a^2V(X)$$

**Theorem 23.**

$$\mu = \Phi'_X(1)$$

$$E(X^2) = \Phi''_X(1) + \Phi'_X(1)$$

## 5.9 Covariance and Correlation

**Theorem 24** (Variance of sum).

$$V(X + Y) = V(X) + V(Y) + 2E[(X - E(X))(Y - E(Y))]$$

**Definition 21** (Covariance). The covariance of  $X, Y$  is given by

$$Cov(X, Y) = E[(X - E(X))(Y - E(Y))] = E(XY) - E(X)E(Y)$$

**Theorem 25.**  $X, Y$  are independent  $\implies Cov(X, Y) = 0$

**Definition 22** (Correlation coefficient). The correlation coefficient of  $X, Y$  is

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{V(X)V(Y)}}$$

## 5.10 Schwartz and Chebyshev's Inequalities

**Theorem 26** (Schwartz inequality).  $[E(XY)]^2 \leq E(X^2)E(Y^2)$

**Corollary 26.1.**  $[Cov(X, Y)]^2 \leq V(X)V(Y)$ . This means  $|\rho(X, Y)| \leq 1$ .

**Theorem 27** (Chebyshev's inequality).

$$P(|X - \mu| \geq t) \leq \frac{\sigma^2}{t^2}, \forall t > 0$$

## 6 Important Discrete Densities

### 6.1 Uniform

$$\begin{aligned}\text{DiscreteUnif}(a, b) \sim f(x) &= \begin{cases} \frac{1}{b-a+1}, & a \leq x \in \mathbb{Z} \leq b \\ 0, & \text{otherwise} \end{cases} \\ \mu &= \frac{a+b}{2} \\ \sigma^2 &= \frac{(a+b-1)^2 - 1}{12}\end{aligned}$$

### 6.2 Bernoulli

$$\begin{aligned}\text{Ber}(p) \sim f(x) &= \begin{cases} p, & x = 1 \\ 1-p, & x = 0 \\ 0, & \text{otherwise} \end{cases} \\ \mu &= p \\ \sigma^2 &= p - p^2\end{aligned}$$

### 6.3 Binomial

$$\begin{aligned}\text{Binom}(n; p) \sim f(x) &= \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & x = 0, 1, \dots, n \\ 0, & \text{otherwise} \end{cases} \\ \mu &= np \\ \sigma^2 &= np(p-1)\end{aligned}$$

### 6.4 Geometric

$$\begin{aligned}\text{Geom}(p) \sim f(x) &= \begin{cases} p(1-p)^x, & x = 0, 1, \dots, n \\ 0, & \text{otherwise} \end{cases} \\ \mu &= \frac{1-p}{p} \\ \sigma^2 &= \frac{1-p}{p^2}\end{aligned}$$

### 6.5 Negative binomial

$$\begin{aligned}\text{NegBinom}(\alpha; p) \sim f(x) &= \begin{cases} \binom{\alpha+x-1}{x} p^\alpha (1-p)^x, & x = 0, 1, \dots, n \\ 0, & \text{otherwise} \end{cases} \\ \mu &= \frac{p\alpha}{1-p} \\ \sigma^2 &= \frac{p\alpha}{(1-p)^2}\end{aligned}$$

### 6.6 Hypergeometric

$$\begin{aligned}\text{HyperGeom}(r, r_1; n) \sim f(x) &= \begin{cases} \frac{\binom{r_1}{x} \binom{r-r_1}{n-x}}{\binom{r}{n}}, & x = 0, 1, \dots, n \\ 0, & \text{otherwise} \end{cases} \\ \mu &= \frac{nr_1}{r} \\ \sigma^2 &= \frac{nr_1(r-r_1)(r-n)}{r^2(r-1)}\end{aligned}$$

Consider a population of  $r$  objects, of which  $r_1$  are of one type and  $r_2 = r - r_1$  are of a second type. Suppose a random sample of size  $n$  is drawn from the population. Let  $X$  be the number of objects of the first type in the sample.

## 6.7 Poisson

$$\begin{aligned}\text{Poisson}(\lambda) \sim f(x) &= \begin{cases} \frac{\lambda^x e^{-\lambda}}{x!}, & x = 0, 1, \dots, n \\ 0, & \text{otherwise} \end{cases} \\ \mu &= \lambda \\ \sigma^2 &= \lambda \\ \Phi(z) &= e^{\lambda(z-1)}\end{aligned}$$

## 6.8 Multinomial

$$\text{Multinom}(n; p_1, p_2, \dots, p_n) \sim f(x_1, x_2, \dots, x_n) \begin{cases} \frac{n!}{x_1! x_2! \dots x_r!} p_1^{x_1} p_2^{x_2} \dots p_n^{x_n}, & x_i \in \mathbb{N} \cup \{0\}, x_1 + x_2 + \dots + x_r = n \\ 0, & \text{otherwise} \end{cases}$$

We have,  $X_i \sim \text{Binom}(n; p_i)$ .

**Theorem 28.** Let  $X_1, X_2, \dots, X_r$  be pair-wise independent. Let  $Y = X_1 + X_2 + \dots + X_r$ .

- $X_i \sim \text{Binom}(n_i; p) \implies Y \sim \text{Binom}(\sum_i n_i; p)$
- $X_i \sim \text{NegBinom}(\alpha_i; p) \implies Y \sim \text{NegBinom}(\sum_i \alpha_i; p)$
- $X_i \sim \text{Poisson}(\lambda) \implies Y \sim \text{Poisson}(\sum_i \lambda_i)$
- $X_i \sim \text{Geom}(p) \implies Y \sim \text{NegBinom}(r; p)$ .

## 7 Continuous Random Variable

**Definition 23** (Continuous random variable). A continuous random variable  $X$  on a probability space  $(\Omega, \mathcal{A}, P)$  is a real-valued function  $X : \Omega \rightarrow \mathbb{R}$  such that for  $x \in \mathbb{R}$ ,  $\{\omega | X(\omega) \leq x\}$  is an event.

$$P(X = x) = 0, \forall x$$

### 7.1 Continuous Distribution

**Definition 24** (Continuous distribution).  $F_X : \mathbb{R} \rightarrow [0, 1]$

$$F_X(x) = P(X \leq x)$$

is called the distribution of continuous random variable  $X$ .

**Theorem 29.** A continuous distribution is right-continuous and non-decreasing.

**Corollary 29.1.** If  $F$  is a continuous distribution,

$$\begin{aligned}\lim_{x \rightarrow -\infty} F(x) &= 0 \\ \lim_{x \rightarrow +\infty} F(x) &= 1\end{aligned}$$

### 7.2 Continuous Density

**Definition 25** (Continuous density function).  $f_X : \mathbb{R} \rightarrow \mathbb{R}$

$$\int_{-\infty}^{\infty} f_X(x) dx = 1$$

$$F(x) = \int_{-\infty}^x f(t) dt, \forall x \in \mathbb{R}$$

If such a density  $f_X$  exists,  $X$  and  $F_X$  are called absolutely continuous.

**Theorem 30.** If  $F_X$  is continuously differentiable (or,  $f$  is continuous) at  $x_0$ ,

$$f(x_0) = F'(x_0)$$



### 7.3 Change of Variable

**Theorem 31** (Change of variable). Let  $Y = \varphi(X)$ , where  $\varphi$  is differentiable and strictly increasing or decreasing in an interval  $I$ . Let  $f_X(x) = 0, \forall x \notin I$ .

$$f_Y(y) = \begin{cases} f(\varphi^{-1}(y)) \left| \frac{d}{dy}(\varphi^{-1}(y)) \right|, & \forall y \in \varphi(I) \\ 0, & \text{otherwise} \end{cases}$$

### 7.4 Joint distribution

**Definition 26** (Joint distribution).  $F : \mathbb{R}^2 \rightarrow [0, 1]$

$$F_{X,Y}(x, y) = P(X \leq x, Y \leq y)$$

**Theorem 32.** Probability in a rectangular region is given by

$$P(a < X \leq b, c < Y \leq d) = F(b, d) - F(a, d) - F(b, c) + F(a, c)$$

### 7.5 Joint Density

**Definition 27** (Joint density).  $f : \mathbb{R}^2 \rightarrow \mathbb{R}$

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \, dx \, dy = 1$$

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(u, v) \, dv \, du$$

**Theorem 33.** If  $f$  is continuous at  $(x_0, y_0)$ ,

$$f(x_0, y_0) = F_{xy}(x_0, y_0)$$

### 7.6 Independence

**Definition 28** (Independence).  $X, Y$  are independent iff

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

or,

$$f(x, y) = f_X(x)f_Y(y), \forall x, y \in \mathbb{R}$$

**Theorem 34.** Let  $X_1, X_2, \dots, X_n$  be pair-wise independent. Suppose,  $Y = \varphi(X_1, X_2, \dots, X_n)$  and  $Z = \psi(X_{m+1}, X_{m+2}, \dots, X_n)$  where  $1 \leq m < n$ .  $Y, Z$  are independent.

### 7.7 Sum

**Theorem 35** (Sum of continuous random variables).

$$F_{X+Y}(z) = \int_{-\infty}^z \int_{-\infty}^{\infty} f_{X,Y}(x, z-x) \, dx \, dv$$

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_{X,Y}(x, z-x) \, dx = f_X(z) * f_Y(z)$$

### 7.8 Conditional Density

**Definition 29** (Conditional density). The conditional density of  $Y$  given  $X$  is defined as

$$f_{Y|X}(y|x) := \begin{cases} \frac{f_{X,Y}(x,y)}{f_X(x)}, & 0 < f_X(x) < \infty \\ 0, & \text{otherwise} \end{cases}$$

$$P(a \leq Y \leq b | X = x) = \int_a^b f_{Y|X}(y|x) \, dy$$

**Theorem 36** (Baye's rule).

$$f_{X|Y}(x|y) = \frac{f_X(x)f_{Y|X}(y|x)}{\int_{-\infty}^{\infty} f_X(x)f_{Y|X}(y|x) \, dx}$$

## 7.9 Expectation

**Definition 30** (Expectation of continuous random variable).

$$E(X) = \int_{-\infty}^{\infty} xf(x) dx$$

**Theorem 37.** Let  $X_1, X_2, \dots, X_n$  have joint density  $f$  and  $Z = \varphi(X_1, \dots, X_n)$ .

$$E(Z) = \int \varphi(x_1, \dots, x_n)f(x_1, \dots, x_n) d\mathbf{x}$$

**Definition 31** (Conditional expectation).

$$E[Y|X=x] = \int_{-\infty}^{\infty} yf(y|x) dy = \frac{\int_{-\infty}^{\infty} yf(x,y) dy}{f_X(x)}$$

is called the conditional expectation of  $Y$  given  $X$  or the regression function of  $Y$  on  $X$ .

## 7.10 Moments

Moments, variance, covariance, correlation are defined like the discrete case.

# 8 Important Continuous Densities

## 8.1 Symmetric

$f$  is symmetric if  $f(-x) = f(x), \forall x$ . If  $X$  and  $-X$  have the same distribution,  $X$  is called a symmetric random variable.

## 8.2 Standard Normal

$$\begin{aligned}\Phi(x) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \\ \mu &= 0 \\ \sigma^2 &= 1\end{aligned}$$

$\Phi$  is symmetric. The density and distribution have the same function.

## 8.3 Normal

For  $\sigma > 0$ ,

$$\begin{aligned}n(\mu, \sigma^2) \sim f(x) &= \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \\ F(x) &= \Phi\left(\frac{x-\mu}{\sigma}\right)\end{aligned}$$

$$X \sim \Phi \implies \mu + \sigma X \sim n(\mu, \sigma^2)$$

## 8.4 Exponential

$$\begin{aligned}\text{Exp}(\lambda) \sim f(x) &= \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases} \\ F(x) &= \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases} \\ \mu &= \frac{1}{\lambda} \\ \sigma^2 &= \frac{1}{\lambda^2}\end{aligned}$$

## 8.5 Gamma

$$\Gamma(\alpha, \lambda) \sim f(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

where  $\Gamma(x) = \int_0^\infty t^{x-1} e^{-t} dt, x > 0$ .

$$\mu = \frac{\alpha}{\lambda}$$

$$\sigma^2 = \frac{\alpha}{\lambda^2}$$

$\text{Exp}(\lambda) \sim \gamma(1, \lambda)$

## 8.6 Cauchy

$$\text{Cauchy}(\alpha, \gamma) \sim f(x) = \frac{1}{\pi \gamma \left[ 1 + \left( \frac{x-\alpha}{\gamma} \right)^2 \right]}$$

$$F(x) = \frac{1}{\pi} \tan^{-1} \left( \frac{x-\alpha}{\gamma} \right) + \frac{1}{2}$$

## 8.7 Standard Bivariate Normal

$$f(x, y) = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}}$$

**Theorem 38.** Let  $X \sim \text{Exp}(\lambda)$  and  $a, b \geq 0$ .

1.  $P(X > a + b) = P(X > a)P(X > b)$
2.  $P(X > a + b | X > a) = P(X > b)$

Both these statements are equivalent. If they hold for some  $X$ ,  $X$  is either exponentially distributed or  $P(X > 0) = 0$ .

**Theorem 39.**  $X \sim n(0, \sigma^2) \implies X^2 \sim \Gamma\left(\frac{1}{2}, \frac{1}{2\sigma^2}\right)$

**Theorem 40.**  $X_i \sim \Gamma(\alpha_i, \lambda) \implies \sum_i X_i \sim \Gamma(\sum_i \alpha_i, \lambda)$

**Theorem 41.**  $X_i \sim n(\mu_i, \sigma_i^2) \implies \sum_i X_i \sim n(\sum_i \mu_i, \sum_i \sigma_i^2)$

**Theorem 42.** Here are some important results that may be useful in computations:

- $\int_{-\infty}^{\infty} e^{-\frac{x^2}{2}} dx = \sqrt{2\pi}$
- $\int_0^\infty x^{\alpha-1} e^{-\lambda x} dx = \frac{\Gamma(\alpha)}{\lambda^\alpha}$
- $\Gamma(x+1) = x\Gamma(x)$
- $\Gamma(1) = 1, \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$
- $\Gamma(n) = (n-1)!, \forall n \in \mathbb{N}$
- $\Gamma\left(\frac{n}{2}\right) = \frac{\sqrt{\pi}(n-1)!}{2^{n-1}\left(\frac{n-1}{2}\right)!}, \forall \text{ odd natural number } n$

## 9 Central Limit Theorem

**Theorem 43** (Central limit theorem). Let  $X_1, X_2, \dots$  be independent, identically distributed random variables with mean  $\mu$  and variance  $\sigma^2$ . Let  $S_n = \sum_i X_i$ .

$$\lim_{n \rightarrow \infty} P\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x\right) = \Phi(x), \forall x \in \mathbb{R}$$

**Corollary 43.1.** For very large  $n$ ,  $\frac{S_n - n\mu}{\sigma\sqrt{n}} \sim \Phi$ .