170E Review Sheet

Theodore Nguyen

UCLA Winter 2022
Professor Harrop-Griffiths

Contents

1 Properties of Probability

1.1 Properties of Probability

Definition 1.1. Probability theory takes place inside a **probability space** $(\Omega, \mathcal{F}, \mathbb{P})$.
This consists of three objects:

1. A non-empty set Ω, called the **sample space**. *(All possible outcomes of an experiment)*

2. A set $\mathcal{F}$ of subsets of Ω satisfying certain properties.

   - Elements of $\mathcal{F}$ are called **events**. *(Outcome of a single experiment)*
   - Events $A_1, A_2, \ldots, A_k$ are called **mutually exclusive** if they are **pairwise disjoint**, i.e.,

   $$\text{If } i \neq j \text{ then } A_i \cap A_j = \emptyset.$$

   - Events $A_1, A_2, \ldots, A_k$ are called **exhaustive** if their union is the sample space, i.e.,

   $$A_1 \cup A_2 \cup \cdots \cup A_k = \bigcup_{j=1}^{k} A_j = \Omega$$

3. A function $\mathbb{P} : \mathcal{F} \to [0, 1]$, called a **probability measure**. This satisfies:

   - $\mathbb{P}[\Omega] = 1$.
   - If $A_1, A_2, \ldots, A_n$ are mutually exclusive events then

   $$\mathbb{P}\left[\bigcup_{j=1}^{n} A_j\right] = \sum_{j=1}^{n} \mathbb{P}[A_j]$$

- If $A_1, A_2, \dots$ are <u>mutually exclusive</u> events then

$$\mathbb{P}\left[\bigcup_{j=1}^{\infty} A_j\right] = \sum_{j=1}^{\infty} \mathbb{P}[A_j]$$

- If A is an event $\mathbb{P}[A]$ is the "probability of A".

Theorem 1.1. $\mathbb{P}[\emptyset] = 0$.

Theorem 1.2. *If $A \subseteq \Omega$ is an event and $A' = \Omega \setminus A$ then*

$$\mathbb{P}[A] = 1 - \mathbb{P}[A']$$

Theorem 1.3. *If $A \subseteq B$ then*

$$\mathbb{P}[B \setminus A] = \mathbb{P}[B] - \mathbb{P}[A]$$

Theorem 1.4. *If $A \subseteq B$ are events then*

$$\mathbb{P}[A] \leq \mathbb{P}[B]$$

2 Properties of Probability

2.1 Properties of Probability

Theorem 2.1. *If $A, B \subseteq \Omega$ are events then*

$$\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B] - \mathbb{P}[A \cap B]$$

Theorem 2.2. *If $A, B, C \subseteq \Omega$ are events then*

$$\begin{aligned}\mathbb{P}[A \cup B \cup C] = \mathbb{P}[A] &+ \mathbb{P}[B] + \mathbb{P}[C] \\ &- \mathbb{P}[A \cap B] - \mathbb{P}[B \cap C] - \mathbb{P}[C \cap A] \\ &+ \mathbb{P}[A \cap B \cap C]\end{aligned}$$

Theorem 2.3. *If $A_1, A_2, \dots, A_n \subseteq \Omega$ are events then*

$$\mathbb{P}\left[\bigcup_{j=1}^{n} A_j\right] \leq \sum_{j=1}^{n} \mathbb{P}[A_j]$$

Remark. This is known as the **union bound**. We can take $n \to \infty$.

2.2 Independence

Definition 2.1.

- We say that two events $A, B \subseteq \Omega$ are **independent** if

$$\mathbb{P}[A \cap B] = \mathbb{P}[A]\mathbb{P}[B]$$

- If two events are not independent we say they are **dependent**.

Definition 2.2. - We say that events $A_1, \dots, A_n \subseteq \Omega$ are mutually independent if, given any $1 \leq k \leq n$ and $1 < j_1 < j_2 < \cdots < j_k \leq n$ we have

$$\mathbb{P}\left[\bigcap_{\ell=1}^{k} A_{j_\ell}\right] = \prod_{\ell=1}^{k} \mathbb{P}[A_{j_\ell}]$$

- In the special case that $n = 3$, this says that events $A, B, C \subseteq \Omega$ are mutually independent if **all** of the following are true:

$$\begin{aligned} \mathbb{P}[A \cap B] &= \mathbb{P}[A]\mathbb{P}[B], \\ \mathbb{P}[B \cap C] &= \mathbb{P}[B]\mathbb{P}[C], \\ \mathbb{P}[C \cap A] &= \mathbb{P}[C]\mathbb{P}[A], \\ \mathbb{P}[A \cap B \cap C] &= \mathbb{P}[A]\mathbb{P}[B]\mathbb{P}[C] \end{aligned}$$

3 Methods of Enumeration

3.1 Methods of Enumeration

Definition 3.1. The probability of an event $A \subseteq \Omega$ is

$$\mathbb{P}[A] = \frac{|A|}{|\Omega|}$$

Definition 3.2 (Multiplication Principle)**.** r **mutually independent** experiments so that:

- The 1^{st} experiment has n_1 possible outcomes.
- The 2^{nd} experiment has n_2 possible outcomes.
- ...
- The r^{th} experiment has n_r possible outcomes.

Then the composite experiment has $n_1 \times n_2 \times \cdots \times n_r$ outcomes.

Theorem 3.1. *There are n^r possible choices of* ***ordered*** *sample of size r from a set of n objects* ***with replacement****.*

Theorem 3.2. *There are*

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

ordered *samples of size r from a set of n objects* ***without replacement****.*

Remark. The number ${}_nP_r$ is known as the number of **permutations** of n objects, taken r at a time.

Theorem 3.3. *There are*

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

unordered *samples of size r from a set of n objects* ***without replacement****.*

Remark. The number ${}_nC_r$ is known as the number of **combinations** of n objects, taken r at a time. Note that ${}_nC_r = {}_nC_{n-r}$.

Theorem 3.4. *There are ${}_{n+r-1}C_r$ possible choice of **unordered** sample of size r from a set of n objects **with replacement**.*

4 Methods of Enumeration

4.1 Methods of Enumeration

Definition 4.1. Given n objects, some are identical. There are ${}_nP_n = n!$ **distinguishable permutations**.

Theorem 4.1. *Suppose I have:*

- n_1 *objects of type 1,*
- n_2 *objects of type 2,*
- ...
- n_r *objects of type r.*

Let $n = n_1 + n_2 + \cdots + n_r$. Then the number of distinguishable permuations is:

$$\binom{n}{n_1, n_2, \ldots, n_r} = \frac{n!}{n_1!n_2!\cdots n_r!}$$

Theorem 4.2. *If $n \geq 0$ then*

$$(x+y)^n = \sum_{r=0}^{n} \binom{n}{r} x^r y^{n-r}$$

*where the **binomial coefficient** is*

$$\binom{n}{r} = {}_nC_r = \binom{n}{r, n-r}$$

Theorem 4.3. *We have $\sum_{r=0}^{n} \binom{n}{r} = 2^n$.*

Theorem 4.4. *If $n, r \geq 0$ then*

$$(x_1 + x_2 + \cdots x_r)^n = \sum_{n_1+\cdots+n_r=n} \binom{n}{n_1, n_2, \ldots, n_r} x_1^{n_1} x_2^{n_2} \cdots x_r^{n_r}$$

5 Conditional Probability

5.1 Conditional Probability

Definition 5.1. $\mathbb{P}[A|B]$ is probability of A conditioned on B.

Definition 5.2. Let $B \subseteq \Omega$ be an event so that $\mathbb{P}[B] \neq 0$. The probability of an event $A \subseteq \Omega$ **conditioned on the event B** is

$$\mathbb{P}[A|B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}$$

Theorem 5.1. *If $B \subseteq \Omega$ is an event so that $\mathbb{P}[B] \neq 0$ then $\mathbb{P}[\cdot|B]$ is a probability measure. Precisely:*

1. $\mathbb{P}[\Omega|B] = 1$.

2. If $A_1, A_2, \ldots, A_n$ *are mutually exclusive events then*

$$\mathbb{P}\left[\bigcup_{j=1}^{n} A_j \middle| B\right] = \sum_{j=1}^{n} \mathbb{P}[A_j|B]$$

3. If $A_1, A_2, \ldots$ *are mutually exclusive events then*

$$\mathbb{P}\left[\bigcup_{j=1}^{\infty} A_j \middle| B\right] = \sum_{j=1}^{\infty} \mathbb{P}[A_j|B]$$

Theorem 5.2. *If A and B are independent events so that $\mathbb{P}[B] \neq 0$ then*

$$\mathbb{P}[A|B] = \mathbb{P}[A]$$

6 Bayes' Theorem

6.1 Conditional Probability

Theorem 6.1 (The Law of Total Probability)**.** *Let $A \subseteq \Omega$ be an event and $B_1, B_2, \ldots, B_n \subseteq \Omega$ be mutually exclusive events so that $\mathbb{P}[B_j] \neq 0$ and*

$$A \subseteq \bigcup_{j=1}^{n} B_j$$

Then

$$\mathbb{P}[A] = \sum_{j=1}^{n} \mathbb{P}[A|B_j]\,\mathbb{P}[B_j]$$

6.2 Bayes' Theorem

Theorem 6.2 (Bayes' Theorem)**.** *If $A, B \subseteq \Omega$ are events so that $\mathbb{P}[A], \mathbb{P}[B] \neq 0$ then*

$$\mathbb{P}[B|A] = \frac{\mathbb{P}[A|B]\,\mathbb{P}[B]}{\mathbb{P}[A]}$$

Theorem 6.3 (Bayes' Theorem (improved version))**.** *If $A \subseteq \Omega$ is an event and $B_1, B_2, \ldots, B_n \subseteq \Omega$ are mutually exclusive events so that $\mathbb{P}[A], \mathbb{P}[B_j] \neq 0$ and*

$$A \subseteq \bigcup_{j=1}^{n} B_j$$

then for any $1 \leq k \leq n$ we have

$$\mathbb{P}[B_k|A] = \frac{\mathbb{P}[A|B_k]\,\mathbb{P}[B_k]}{\sum_{j=1}^{n} \mathbb{P}[A|B_j]\,\mathbb{P}[B_j]}$$

7 Discrete Random Variables

7.1 Discrete Random Variables

Definition 7.1.

- Given a set S, a **random variable** is a function $X : \Omega \to S$ satisfying certain properties.
- For the sake of this class, we can assume that all functions $X : \Omega \to S$ are random variables.
- Notation: if $x \in S$ and $A \subseteq S$

$$\mathbb{P}[X = x] = \mathbb{P}[\{\omega \in \Omega : X(\omega) = x\}]$$
$$\mathbb{P}[X \in A] = \mathbb{P}[\{\omega \in \Omega : X(\omega) \in A\}]$$

Definition 7.2.

- Let $X : \Omega \to S$ be a random variable.
- We say that X is a **discrete random variable** if $S \subseteq \mathbb{R}$ is a countable (i.e. finite or in one-to-one correspondence with $\mathbb{N}$) set.

Definition 7.3. Let X be a discrete random variable taking values in a countable set $S \subseteq \mathbb{R}$.

- We define the **probability mass function** (PMF) of X to be the function $p_X : S \to [0, 1]$ given by

$$p_X(x) = \mathbb{P}[X = x]$$

- We define the **cumulative distribution function** (CDF) of X to be the function $F_X : \mathbb{R} \to [0, 1]$ given by

$$F_X(x) = \mathbb{P}[X \leq x]$$

- We say that two random variables X, Y are **identically distributed** if they have the same CDF and write $X \sim Y$.

Important Example 1

Uniform Distribution

- Let $m \geq 1$.
- We say that a discrete random variable X is **uniformly distributed** on $\{1, 2, \ldots, m\}$ and write $X \sim \text{Uniform}(\{1, 2, \ldots, m\})$ if it has PMF

$$\mathbb{P}[X = x] = p_X(x) = \frac{1}{m} \text{ if } x \in \{1, 2, 3, \ldots, m\}$$

- If $X \sim \text{Uniform}(\{1, 2, \ldots, m\})$ then it has CDF

$$F_X(x) = \begin{cases} 0 & \text{if } x < 1 \\ \frac{k}{m} & \text{if } x < k + 1 \quad \text{and} \quad k \in \{1, 2, \ldots, m - 1\} \\ 1 & \text{if} \quad x \geq m \end{cases}$$

Theorem 7.1. *If X is a discrete random variable taking values in a countable set $S \subseteq \mathbb{R}$ and $A \subseteq \mathbb{R}$ is any set then*

$$\mathbb{P}[X \in A] = \sum_{x \in A \cap S} p_X(x)$$

Theorem 7.2. *If X is a discrete random variable taking values in a countable set $S \subseteq \mathbb{R}$ then*

$$F_X(x) = \sum_{\substack{y \leq x \\ y \in S}} p_X(y)$$

Theorem 7.3. *If X is a discrete random variable and $a < b$ then*

$$\mathbb{P}[a < X \leq b] = F_X(b) - F_X(a)$$

Theorem 7.4. *If X is a discrete random variable taking values in a countable set $S \subseteq \mathbb{R}$ then*

$$\sum_{x \in S} p_X(x) = 1$$

Theorem 7.5. *If X is a discrete random variable then:*

- *F_x is non-decreasing and right-continuous.*
- $\lim_{x \to -\infty} F_X(x) = 0$
- $\lim_{x \to +\infty} F_X(x) = 1$

8 Mathematical Expectation

8.1 Mathematical Expectation

Definition 8.1. If X is a discrete random variable taking values in a countable set $S \subseteq \mathbb{R}$, we define its **expected value** to be

$$\mathbb{E}[X] = \sum_{x \in S} x \, p_X(x)$$

provided the sum converges in a suitable sense.

Remark. We often use the notation $\mu_X = \mathbb{E}[X]$ *("Mean" or "Average" Value)*

Important Example 2

Bernoulli Random Variable

- Let $p \in (0, 1)$.
- We say that discrete random variable X is a **Bernoulli random variable** and write $X \sim$ Bernoulli(p) if it has PMF

$$p_X(x) = \begin{cases} p & \text{if} \quad x = 1 \\ 1 - p & \text{if} \quad x = 0 \end{cases}$$

- $\mathbb{E}[X] = p$

Theorem 8.1. *If $X \sim$ Uniform$(\{1, 2, \ldots, m\}$ then*

$$\mathbb{E}[X] = \frac{m+1}{2}$$

Definition 8.2. If X is a discrete random variable taking values in a countable set $S \subseteq \mathbb{R}$ and $g : S \to \mathbb{R}$ is a function, we define the **expected value of** $g(X)$ to be

$$\mathbb{E}[g(x)] = \sum_{x \in S} g(x) \, p_X(x)$$

provided the sum converges in a suitable sense.

Theorem 8.2. *Let X be a discrete random variable. If $a \in R$ then $\mathbb{E}[a] = a$.*

Theorem 8.3. *Let X be a discrete random variable taking values in a countable set $S \in \mathbb{R}$. If $a, b \in \mathbb{R}$ and $g, h : S \to \mathbb{R}$ then*

$$\mathbb{E}[a \, g(X) + b \, h(X)] = a \, \mathbb{E}[g(X)] + b \, \mathbb{E}[h(x)]$$

Theorem 8.4. *Let X be a discrete random variable taking values in a countable set $S \subseteq \mathbb{R}$. If $g, h : S \to \mathbb{R}$ satisfy $g(x) \leq h(x)$ for all $x \in S$ then*

$$\mathbb{E}[g(X)] \leq \mathbb{E}[h(X)]$$

9 Special Mathematical Expectations

9.1 Special Mathematical Expectations

Definition 9.1. Let X be a discrete random variable taking values in a discrete set $S \subseteq \mathbb{R}$ and $b \in \mathbb{R}$. We define **the r^{th} moment of X about b** to be

$$\mathbb{E}[(X-b)^r] = \sum_{x \in S} (x-b)^r p_X(x)$$

Remark. When $b = 0$ we refer to this as simply the **r^{th} moment** of X. ($\mathbb{E}[X^r]$)

Definition 9.2. Let X be a discrete random variable. We define the **variance** of X to be

$$\operatorname{var}(X) = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right]$$

whenever it converges. We use the notation $\sigma_X^2 = \operatorname{var}(X)$.

Remark. The **standard deviation** of X is $\sigma_X = \sqrt{\operatorname{var}(X)}$.

Theorem 9.1. *If X is a discrete random variable and $a, b \in \mathbb{R}$ then:*

- $\mathbb{E}[a\,X + b] = a\,\mathbb{E}[X] + b$
- $var(a\,X + b) = a^2\,var(X)$

Theorem 9.2. *If X is a discrete random variable then*

$$var(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

Definition 9.3. If X is a discrete random variable we define the **Moment Generating Function** (MGF) of X to be the function

$$M_X(t) = \mathbb{E}[e^{tX}]$$

whenever it exists.

Theorem 9.3. *Let X be a discrete random variable with MGF $M_X(t)$ that is well-defined and smooth for $t \in (-\delta, \delta)$. Then:*

$$\left.\frac{d^r}{dt^r} M_X\right|_{t=0} = \mathbb{E}[X^r]$$

Theorem 9.4. *Let X be a discrete random variable with MGF $M_X(t)$ that is well-defined and smooth for $t \in (-\delta, \delta)$.*

- $\left.\dfrac{d}{dt} \ln M_X\right|_{t=0} = \mathbb{E}[X]$
- $\left.\dfrac{d^2}{dt^2} \ln M_X\right|_{t=0} = var[X]$

10 The Binomial and Geometric Distributions

10.1 The Binomial Distribution

Important Example 3

Binomial Distribution

- A **Bernoulli trial** is an experiment that has probability $p \in (0, 1)$ of success and probability $(1 - p)$ of failure.
- Suppose we run $n \geq 1$ independent, identical Bernoulli trials.
- Let X be the number of successes.
- X is a discrete random variable taking values in the set $S = \{0, 1, \ldots, n\}$.
- We say that X is a **Binomial random variable** with parameters n, p and write $X \sim \text{Binomial}(n, p)$.

Theorem 10.1. *If $X \sim Binomial(n, p)$ then its PMF is*

$$p_X(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad \text{if} \quad x \in \{0, 1, \ldots, n\}$$

Theorem 10.2. *If $X \sim Binomial(n, p)$ its MGF is*

$$M_X(t) = (1 - p + pe^t)^n$$

Theorem 10.3. *If $X \sim Binomial(n, p)$ its mean is*

$$\mathbb{E}[X] = np$$

Theorem 10.4. *If $X \sim Binomial(n, p)$ its variance is*

$$var(X) = np(1-p)$$

10.2 The Geometric Distribution

Important Example 4

Geometric Distribution

- Suppose we repeatedly run independent, identical Bernoulli trials with probability $p \in (0, 1$ of success.
- Let X be the trial on which we first achieve success.
- X is a discrete random variable taking values in the set $S = \{1, 2, 3, \dots\}$.
- We say that X is a **geometric random variable** with parameter p and write $X \sim$ Geometric(p).

Theorem 10.5. *If $X \sim$ Geometric(p) then its PMF is*

$$p_X(x) = (1-p)^{x-1}p \quad \textit{if} \quad x \in \{1, 2, 3, \dots\}$$

Theorem 10.6. *If $X \sim$ Geometric(p) then its MGF is*

$$M_X(t) = \frac{e^t p}{1-(1-p)e^t} \quad \textit{if} \quad t < -\ln(1-p)$$

Theorem 10.7. *If $X \sim$ Geometric(p) then its mean is*

$$\mathbb{E}[x] = \frac{1}{p}$$

Theorem 10.8. *If $X \sim$ Geometric(p) then its variance is*

$$var(X) = \frac{1-p}{p^2}$$

Remark (Hypergeometric Distribution (2.5 pg.71)).

$$f(x) = \mathbb{P}(X = x) = \frac{\binom{N_1}{x}\binom{N_2}{n-x}}{\binom{N}{n}}$$

11 The Negative Binomial and Poisson Distributions

11.1 The Negative Binomial Distribution

Important Example 5

Negative Binomial Distribution

- Suppose we repeatedly run independent, identical Bernoulli trials with probability $p \in (0,1)$ of success.
- Let $r \geq 1$ and let X be the trial on which we first achieve the r^{th} success.
- X is a discrete random variable taking values in the set $S = \{r, r+1, r+2, \ldots\}$.
- We say that X is a **negative binomial random variable** with parameters r, p and write $X \sim \text{Negative Binomial}(r,p)$.

Remark. Negative Binomial$(1,p) \sim$ Geometric(p)

Theorem 11.1. *If $X \sim$ Negative Binomial(r,p) then its PMF is*

$$p_X(x) = \binom{x-1}{r-1} p^r (1-p)^{x-r} \quad \text{if} \quad x \in \{r, r+1, \ldots\}$$

Theorem 11.2. *If $r \geq 1$ is an integer and $0 < s < 1$ then*

$$\left(\frac{1}{1-s}\right)^r = \sum_{x=r}^{\infty} \binom{x-1}{r-1} s^{x-r}$$

Theorem 11.3. *If $X \sim$ Negative Binomial(r,p) then its MGF is*

$$M_X(t) = \left(\frac{e^t p}{1-(1-p)e^t}\right)^r \quad \text{if} \quad t < -\ln(1-p)$$

Theorem 11.4. *If $X \sim$ Negative Binomial(r,p) then*

$$\mathbb{E}[X] = \frac{r}{p}$$
$$var(X) = \frac{r(1-p)}{p^2}$$

11.2 The Poisson Distribution

Important Example 6

Poisson Random Variable

- We make the following assumptions about the arrivals:
  (1) If the time intervals $(a_1, b_1], (a_2, b_2], \ldots, (a_n, b_n]$ are disjoint then the number of arrivals in each time interval are independent.
  (2) If $h = b - a$ is sufficiently small then the probability of exactly one interval $(a, b]$ is λh.
  (3) If $h = b - a$ then the probability of having more than one arrival in the time interval $(a, b]$ converges to zero as $h \to 0$.
- An arrival process satisfying these assumptions is called an **approximate Poisson process**.
- Take X to be the number of arrivals in a unit of time. Then X is called a **Poisson random variable** and we write $X \sim \text{Poisson}(\lambda)$.

Theorem 11.5. *If $X \sim Poisson(\lambda)$ then it has PMF*

$$p_X(x) = e^{-\lambda}\frac{\lambda^x}{x!} \quad \text{if} \quad x \in \{0, 1, 2, \ldots\}$$

12 Random Variables of the Continuous Type

12.1 The Poisson Distribution

Theorem 12.1. *Consider an approximate Poisson process with rate $\lambda > 0$ per unit time. Let X be the number of arrivals in a time interval of length $T > 0$ units. Then $X \sim Poisson(\lambda T)$.*

Theorem 12.2. *If $\lambda > 0$ and $X \sim Poisson(\lambda)$ then its MGF is*

$$M_X(t) = e^{\lambda(e^t - 1)}$$

Theorem 12.3. *If $\lambda > 0$ and $X \sim Poisson(\lambda)$ then*

$$\mathbb{E}[X] = \lambda$$
$$var(X) = \lambda$$

12.2 Random Variables of the Continuous Type

Definition 12.1.

- Let $X : \Omega \to \mathbb{R}$ be a random variable
- We say that X is a **continuous random variable** if there exists a non-negative integrable function $f_X : \mathbb{R} \to [0, \infty)$ so that

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\, dt$$

Note that this ensures that $\underline{F_X(x) \text{ is continuous}}$.

- We call $f_X(x)$ a **probability density function** for X.

Theorem 12.4. *If X is a continuous random variable with PDF $f_X : \mathbb{R} \to [0, \infty)$ then*

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

Theorem 12.5. *If X is a continuous random variable with PDF $f_X : \mathbb{R} \to [0, \infty)$ and $a < b$ then*

$$\mathbb{P}[a < X \leq b] = \int_a^b f_X(x)\, dx$$

Theorem 12.6. *If X is a continuous random variable with PDF $f_X : \mathbb{R} \to [0, \infty)$ then for all $x \in \mathbb{R}$ we have*

$$\mathbb{P}[X = x] = 0$$

13 Random Variables of the Continuous Type

13.1 Random Variables of the Continuous Type

Important Example 7

Uniform Distribution

- Let $a < b$.
- Pick a point X at a random in the interval $[a, b]$.
- If there is an equal probability of picking every point in $[a, b]$, we say that X is **uniformly distributed on the interval** $[a, b]$.
- We write $X \sim \text{Uniform}([a, b])$.

Theorem 13.1. *If $a < b$ and $X \sim$ Uniform$([a, b])$ then it has PDF*

$$f_X(x) = \begin{cases} \frac{1}{b-a} & \textit{if} \quad x \in (a, b) \\ 0 & \textit{otherwise} \end{cases}$$

Theorem 13.2. *We have*

$$\mathbb{E}[X_n] \to \int_{-\infty}^{\infty} x f_X(x)\, dx \quad \textit{as } n \to \infty$$

Definition 13.1.

- If X is a continuous random variable with PDF $f_X(x)$ we define its **expected value** to be

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

- We still use the notation $\mu_X = \mathbb{E}[X]$

- More generally, if $g : \mathbb{R} \to \mathbb{R}$ is any function, we define

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx$$

Theorem 13.3. *Let X be a continuous random variable.*

- *If $a \in \mathbb{R}$ is a constant then*

$$\mathbb{E}[a] = a$$

- *If $a, b \in \mathbb{R}$ are constants and $g, h : \mathbb{R} \to \mathbb{R}$ then*

$$\mathbb{E}[a\, g(X) + b\, h(X)] = a\, \mathbb{E}[g(X)] + b\, \mathbb{E}[h(x)]$$

- *If $g(x) \le h(x)$ for all $x \in \mathbb{R}$ then*

$$\mathbb{E}[g(X)] \le \mathbb{E}[h(X)]$$

14 The Exponential Distribution

14.1 Random Variables of the Continuous Type

Definition 14.1.

- If X is a continuous random variable we define its **variance** to be

$$\mathrm{var}(X) = \mathbb{E}\Big[(X - \mathbb{E}[X])^2\Big]$$

- We still use the notation $\sigma_X^2 = \mathrm{var}(X)$ and define the **standard deviation** to be $\sigma_X = \sqrt{\mathrm{var}(X)}$.

Theorem 14.1. *If X is a continuous random variable then*

$$var(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

Theorem 14.2. *Let $a < b$ and $X \sim Uniform[a, b]$. Then*

$$var(X) = \frac{(b-a)^2}{12}$$

Definition 14.2. If X is a continuous random variable we define its **moment generating function** to be

$$M_X(t) = \mathbb{E}[e^{tX}]$$

for all $t \in \mathbb{R}$ for which this makes sense.

Theorem 14.3. *If X is a continuous random variable with MGF $M_X(t)$ that is smooth on some interval $(-\delta, \delta)$ then for all $n \ge 0$*

$$\frac{d^n}{dt^n} M_X \bigg|_{t=0} = \mathbb{E}[X^n]$$

Theorem 14.4. *If X is a continuous random variable with MGF $M_X(t)$ that is smooth on some interval $(-\delta, \delta)$ then*

$$\frac{d}{dt} \ln M_X \bigg|_{t=0} = \mathbb{E}[X]$$
$$\frac{d^2}{dt^2} \ln M_X \bigg|_{t=0} = var(X)$$

14.2 The Exponential Distribution

Important Example 8

Exponential Distribution

- Consider an approximate Poisson process with rate $\lambda > 0$ per unit time.
- Let X be the time of the first arrival.
- We say that X is **exponentially distributed** with mean waiting time $\theta = \frac{1}{\lambda}$ and write $X \sim \text{Exponential}(\theta)$

Theorem 14.5. *If $\theta > 0$ and $X \sim$ Exponential(θ) then its PDF is*

$$f_X(x) = \frac{1}{\theta} e^{-\frac{x}{\theta}} \quad \textit{if} \quad x > 0$$

Theorem 14.6. *If $\theta > 0$ and $X \sim$ Exponential(θ) then its MGF is*

$$M_X(t) = \frac{1}{1 - \theta t} \quad \textit{if} \quad t < \frac{1}{\theta}$$

Theorem 14.7. *If $\theta > 0$ and $X \sim$ Exponential(θ) then*

$$\mathbb{E}[X] = \theta$$
$$var(X) = \theta^2$$

15 The Gamma Distribution

15.1 The Gamma Distribution

Important Example 9

Gamma Distribution

- Consider an approximate Poisson process with rate $\lambda > 0$ per unit time.
- Let $\alpha \geq 1$ be an integer and X be the time of the α^{th} arrival.
- We say that X is **gamma distributed** with parameters $\alpha, \theta = \frac{1}{\lambda}$ and write $X \sim \text{Gamma}(\alpha, \theta)$

Theorem 15.1. *Let $\alpha \geq 1$ be an integer and $\theta > 0$. If $X \sim Gamma(\alpha, \theta)$ then its PDF is*

$$f_X(x) = \frac{1}{\theta^\alpha (\alpha - 1)!} x^{\alpha - 1} e^{-\frac{x}{\theta}} \quad \textit{if} \quad x > 0$$

Definition 15.1. For $\alpha > 0$ we define the gamma function

$$\Gamma(\alpha) = \int_0^\infty x^{a-1} e^{-x}\, dx$$

Theorem 15.2.

- $\Gamma(1) = 1$
- *For $\alpha > 1$ we have*
$$\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$$
- *If $\alpha \geq 1$ is an integer then*
$$\Gamma(a) = (\alpha - 1)!$$

Theorem 15.3. *$\forall \alpha, \theta > 0$ and $X \sim Gamma(\alpha, \theta)$, then PDF is:*

$$f_X(x) = \frac{1}{\theta^\alpha \Gamma(\alpha)} x^{\alpha - 1} e^{-\frac{x}{\theta}} \quad \textit{if} \quad x > 0$$

Theorem 15.4. *Let $\alpha, \theta > 0$ and $X \sim Gamma(\alpha, \theta)$. Then X has MGF*

$$M_X(t) = \frac{1}{(1 - \theta t)^\alpha} \quad \textit{if} \quad t < \frac{1}{\theta}$$

Theorem 15.5. *Let $\alpha, \theta > 0$ and $X \sim Gamma(\alpha, \theta)$. Then*

$$\mathbb{E}[X] = \alpha\theta$$
$$var(X) = \alpha\theta^2$$

Process	Bernoulli	Poisson
Number of arrivals	Binomial	Poisson
Time of 1^{st} arrival	Geometric	Exponential
Time of k^{th} arrival	Negative Binomial	Gamma

Definition 15.2 (A Special Case)**.**

- If $r \in \{1, 2, 3, \ldots, \}$ we call the $\Gamma(\frac{r}{2}, 2)$ distribution the **chi-square distribution** with r degrees of freedom.
- If $X \sim \chi^2(r)$ then it has PDF

$$f_X(x) = \frac{1}{2^{r/2}\Gamma(r/2)} x^{r/2 - 1} e^{-x/2} \quad \text{if} \quad x > 0$$

as well as

$$\mathbb{E}[X] = r \quad \text{and} \quad \text{var}(X) = 2r$$

Remark (Important Fact). Suppose that X is a continuous random variable with PDF

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

Then $X^2 \sim \chi^2(1)$.

16 The Normal Distribution

16.1 The Normal Distribution

Important Example 10

Normal Distribution

- We say a continuous random variable X is **normally distributed** with mean μ and variance σ^2 if it has PDF
$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{for} \quad x \in \mathbb{R}$$
- We write $X \sim \mathcal{N}(\mu, \sigma^2)$.
- If $\mu = 0$ and $\sigma^2 = 1$ we say that X is a **standard normal** random variable.

Theorem 16.1. *We have*

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(t-\mu)^2}{2\sigma^2}} \, dt = 1$$

Theorem 16.2. *If $X \sim \mathcal{N}(\mu, \sigma^2)$ then it has MGF*

$$M_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}$$

Theorem 16.3. *If $X \sim \mathcal{N}(\mu, \sigma^2)$ then*

$$\mathbb{E}[X] = \mu$$
$$var(X) = \sigma^2$$

Definition 16.1.

$$\phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \, dt = \mathbb{P}[X \leq x] \quad \text{if} \quad X \sim \mathcal{N}(0,1)$$

Theorem 16.4. *If*

$$\phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \, dt$$

is the CDF of the standard normal distribution then

$$\phi(-x) = 1 - \phi(x)$$

Theorem 16.5. *If* $X \sim \mathcal{N}(0,1)$*, then* $-X \sim \mathcal{N}(0,1)$*.*

Theorem 16.6. *If* $X \sim \mathcal{N}(\mu, \sigma^2)$ *then* $Z = \frac{1}{\sigma}(X - \mu) \sim \mathcal{N}(0,1)$*.*

Theorem 16.7. *If* $X \sim \mathcal{N}(\mu, \sigma^2)$ *then*

$$F_X(x) = \mathbb{P}[X \leq x] = \phi\left(\frac{x - \mu}{\sigma}\right)$$

17 Bivariate Distributions of the Discrete Type

17.1 Bivariate Distributions of the Discrete Type

Definition 17.1.

- Let X, Y be a pair of discrete random variables taking values in sets $S_X, S_Y \subseteq \mathbb{R}$.
- We think of (X, Y) as being a random point in $\mathbb{R}^2$ taking values in the set

$$S = S_X \times S_Y = \{ (x, y) : x \in S_X \text{ and } y \in S_y \}$$

- We define the **joint probability mass function** of X, Y to be the function $p_{X,Y}\colon S \to [0,1]$ by

$$p_{X,Y}(x, y) = \mathbb{P}[X = x, Y = y] = \mathbb{P}[(X, Y) = (x, y)]$$

Theorem 17.1. *Let* X, Y *be discrete random variables taking values in sets* $S_X, S_Y \subseteq \mathbb{R}$ *and let* $S = S_X \times S_Y$*. If* X, Y *have joint PMF* $P_{X,Y}(x, y)$ *and* $A \subseteq \mathbb{R}^2$ *then*

$$\mathbb{P}[(X, Y) \in A] = \sum_{(x,y) \in A \cap S} p_{X,Y}(x, y)$$

Theorem 17.2. *Let* X, Y *be discrete random variables taking values in sets* $S_X, S_Y \subseteq \mathbb{R}$ *and let* $S = S_X \times S_Y$*. If* X, Y *have joint PMF* $p_{X,Y}(x, y)$ *then*

$$\sum_{(x,y) \in S} p_{X,Y}(x, y) = 1$$

Definition 17.2.

- Let X, Y be discrete random variables taking values in sets $S_X, S_Y \subseteq \mathbb{R}$.
- We define the **marginal probability mass function of X** to be the function $p_X : S_X \to [0,1]$ given by

$$p_X(x) = \mathbb{P}[X = x]$$

- We define the **marginal probability mass function of Y** to be the function $p_Y : S_Y \to [0,1]$ given by

$$p_Y(y) = \mathbb{P}[Y = y]$$

Theorem 17.3. *Let X, Y be discrete random variables taking values in sets $S_X, S_Y \subseteq \mathbb{R}$. Let X, Y have joint PMF $p_{X,Y}(x, y)$. Then,*

$$p_X(x) = \sum_{y \in S_y} p_{X,Y}(x, y) \quad \textit{and} \quad p_Y(y) = \sum_{x \in S_x} p_{X,Y}(x, y)$$

Theorem 17.4. *Let X, Y be discrete random variables taking values in sets $S_X, S_Y \subseteq \mathbb{R}$. Let X have marginal PMF $p_X(x)$ and Y have marginal PMF $p_Y(y)$. Then,*

$$\sum_{x \in S_x} p_X(x) = 1 \quad \textit{and} \quad \sum_{y \in S_y} p_Y(y) = 1$$

Definition 17.3.

- Let X, Y be discrete random variables taking values in sets $S_X, S_Y \subseteq \mathbb{R}$ and let $S = S_X \times S_Y$.
- We say that random variables X, Y are **independent** if the events $\{X = x\}$ and $\{Y = y\}$ are independent for all $(x, y) \in s$.
- Equivalently, we have

$$p_{X,Y}(x, y) = p_X(x)\, p_Y(y) \quad \text{for all} \quad (x, y) \in S$$

18 Bivariate Distributions of the Discrete Type

18.1 Bivariate Distributions of the Discrete Type

Definition 18.1.

- Let X, Y be a pair of discrete random variables taking values in sets $S_X, S_Y \subseteq \mathbb{R}$ and let $S = S_X \times S_Y$.
- Let X, Y have joint PMF $p_{X,Y}(x, y)$.
- If $g : S \to \mathbb{R}$ we define the **expected value of $g(X, Y)$** to be

$$\mathbb{E}[g(X, Y)] = \sum_{(x,y) \in S} g(x, y)\, p_{X,Y}(x, y)$$

Theorem 18.1. *Let X, Y be discrete random variables taking values in sets $S_X, S_Y \subseteq \mathbb{R}$ and let $S = S_X \times S_Y$.*

- *If $a, b \in \mathbb{R}$ are constants and $g, h : S \to \mathbb{R}$ then*

$$\mathbb{E}[a\, g(X, Y) + b\, h(X, Y)] = a\, \mathbb{E}[g(X, Y)] + b\, \mathbb{E}[h(X, Y)]$$

- *If $g(x, y) \leq h(x, y)$ for all $(x, y) \in S$ then*

$$\mathbb{E}[g(X, Y)] \leq \mathbb{E}[h(X, Y)]$$

- *If $a \in \mathbb{R}$ is a constant, $\mathbb{E}[a] = a$.*

Theorem 18.2. *Let X, Y be discrete random variables taking values in sets $S_X, S_Y \subseteq \mathbb{R}$. Let $g : S_X \to \mathbb{R}$ and $h : S_Y \to \mathbb{R}$. Then*

$$\mathbb{E}[g(X)] = \sum_{x \in S_X} g(x)\, p_X(x) \quad \text{and} \quad \mathbb{E}[h(Y)] = \sum_{y \in S_Y} h(y)\, p_Y(y)$$

Theorem 18.3. *Let X, Y be* ***independent*** *discrete random variables taking values in sets $S_X, S_Y \subseteq \mathbb{R}$. Let $g : S_X \to \mathbb{R}$ and $h : S_Y \to \mathbb{R}$. Then*

$$\mathbb{E}[g(X)\, h(Y)] = \mathbb{E}[g(X)]\, \mathbb{E}[h(Y)]$$

Theorem 18.4 (The Cauchy-Schwarz Inequality)**.** *Let X, Y be discrete random variables. Then*

$$\Big|\mathbb{E}[XY]\Big| \leq \sqrt{\mathbb{E}[X^2]\mathbb{E}[Y^2]}$$

19 The Correlation Coefficient

19.1 The Correlation Coefficient

Definition 19.1.

- Let X, Y be a pair of (discrete) random variables.
- We define the **covariance** of X, Y to be

$$\text{cov}(X, Y) = \mathbb{E}\big[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\big]$$

- We use the notation $\sigma_{XY} = \text{cov}(X, Y)$.

Theorem 19.1. *If X, Y are random variables then*

$$cov(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X]\, \mathbb{E}[Y]$$

Theorem 19.2. *If X is a random variable then*

$$cov(X, X) = var(X)$$

Theorem 19.3. *Let X, Y be independent discrete random variables. Then*

$$cov(X, Y) = 0$$

Theorem 19.4. *If X, Y are (discrete) random variables and $a, b \in \mathbb{R}$ then*

$$cov(aX, bY) = ab\, cov(X, Y)$$

Definition 19.2.

- Let X, Y be a pair of (discrete) random variables.

- We define the **correlation coefficient** of X, Y to be

$$\rho(X,Y) = \frac{\text{cov}(X,Y)}{\sqrt{\text{var}(X)\text{var}(Y)}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

Theorem 19.5. *If X, Y are (discrete) random variables and $a, b > 0$ then*

$$\rho(aX, bY) = \rho(X,Y)$$

Theorem 19.6. *If X, Y are (discrete) random variables then*

$$-1 \leq \rho(X,Y) \leq 1$$

20 Conditional Distributions

20.1 Conditional Distributions

Definition 20.1.

- Let X, Y be a pair of discrete random variables taking values in sets $S_X, S_Y \subseteq \mathbb{R}$.
- If $y \in S_Y$, we define the **random variable $X|y$** with PMF

$$\begin{aligned} p_{X|Y}(x|y) &= \mathbb{P}[X = x | Y = y] \quad \text{for} \quad x \in S_X \\ &= \frac{p_{X,Y}(x,y)}{p_Y(y)} \quad [\text{provided } p_Y(y) \neq 0] \end{aligned}$$

- If $x \in S_X$, we define the **random variable $Y|x$** with PMF

$$\begin{aligned} p_{Y|X}(y|x) &= \mathbb{P}[Y = y | X = x] \quad \text{for} \quad y \in S_Y \\ &= \frac{p_{X,Y}(x,y)}{p_X(x)} \quad [\text{provided } p_X(x) \neq 0] \end{aligned}$$

Theorem 20.1. *Let X, Y be discrete random variables taking values in sets $S_X, S_Y \subseteq \mathbb{R}$.*

- *Given $y \in S_Y$ we have*
$$\sum_{x \in S_x} p_{X|Y}(x|y) = 1$$
- *Given $x \in S_X$ we have*
$$\sum_{y \in S_y} p_{Y|X}(y|x) = 1$$

Definition 20.2.

- Let X, Y be discrete random variables taking values in sets $S_X, S_Y \subseteq \mathbb{R}$.
- Define the function $g : S_X \to \mathbb{R}$ by
$$g(x) = \mathbb{E}[Y|x]$$

- We define the **conditional expectation** of Y conditioned on X to be the **random variable**

$$\mathbb{E}[Y|X] = g(X)$$

- Can similarly define

$$\mathbb{E}[X|Y]$$

Theorem 20.2 (The Law of Iterated Expectation). *Let X, Y be discrete random variables. Then*

$$\mathbb{E}\big[\mathbb{E}[Y|X]\big] = \mathbb{E}[Y]$$

Definition 20.3.

- Let X, Y be discrete random variables taking values in sets $S_X, S_Y \subseteq \mathbb{R}$.
- Define the function $h : S_X \to \mathbb{R}$ by

$$h(x) = \mathrm{var}(Y|x)$$

- We define the **conditional variance** of Y conditioned on X to be the **random variable**

$$\mathrm{var}(Y|X) = h(X)$$

- Can similarly define $\mathrm{var}(X|Y)$

Theorem 20.3 (The Law of Total Variance). *Let X, Y be discrete random variables. Then*

$$\mathbb{E}\big[var(Y|X)\big] + var\big(\mathbb{E}[Y|X]\big) = var(Y)$$

21 Bivariate Distributions of the Continuous Type

21.1 Bivariate Distributions of the Continuous Type

Definition 21.1. Given two continuous random variables X, Y we may define a **joint probability density function**

$$f_{X,Y} : \mathbb{R}^2 \to [0, \infty)$$

so that for any $A \subseteq \mathbb{R}^2$ we have

$$\mathbb{P}[(X, Y) \in A] = \iint_A f_{X,Y}(x, y)\, dxdy$$

Theorem 21.1. *If X, Y are continuous random variables with joint PDF $f_{X,Y}(x, y)$ then*

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

Definition 21.2. Let X, Y be continuous random variables with joint PDF $f_{X,Y}(x, y)$.

- We define the **marginal PDF of X** to be

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

- We define the **marginal PDF of Y** to be

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

Theorem 21.2. *Let X, Y be continuous random variable and f_X be the marginal PDF of X. If $a < b$ then*

$$\mathbb{P}[a < X \leq b] = \int_a^b f_X(x)\,dx$$

Definition 21.3. Let X, Y be continuous random variables with joint PDF $f_{X,Y}(x,y)$. Given a function $g : \mathbb{R}^2 \to \mathbb{R}$ we define the **expected value** of $g(X,Y)$ to be

$$\mathbb{E}[g(X,Y)] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)\,f_{X,Y}(x,y)\,dxdy$$

Theorem 21.3. *Let X, Y be continuous random variables taking values. Then:*

- *If $a \in \mathbb{R}$ is a constant,*

$$\mathbb{E}[a] = a$$

- *If $a, b \in \mathbb{R}$ and $g, h : \mathbb{R}^2 \to \mathbb{R}$ then*

$$\mathbb{E}[a\,g(X,Y) + b\,h(X,Y)] = a\,\mathbb{E}[g(X,Y)] + b\,\mathbb{E}[h(X,Y)]$$

- *If $g, h : \mathbb{R}^2 \to \mathbb{R}$ and $g(x,y) \leq h(x,y)$ for all $(x,y) \in \mathbb{R}^2$ then*

$$\mathbb{E}[g(X,Y)] \leq \mathbb{E}[h(X,Y)]$$

Theorem 21.4. *Let X, Y be continuous random variables with joint PDF $f_{X,Y}(x,y)$ and marginal PDFs $f_X(x)$, $f_Y(y)$. Then if $g, h : \mathbb{R} \to \mathbb{R}$ we have*

$$\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x)\,f_X(x)\,dx$$
$$\mathbb{E}[h(Y)] = \int_{-\infty}^{\infty} h(y)\,f_Y(y)\,dy$$

Definition 21.4. Let X, Y be continuous random variables with joint PDF $f_{X,Y}(x,y)$ and marginal PDFs $f_X(x)$, $f_Y(y)$. We say that X, Y are **independent** if

$$f_{X,Y}(x,y) = f_X(x)\,f_Y(y)$$

for all $(x,y) \in \mathbb{R}^2$.

Theorem 21.5. *Let X, Y be* ***independent*** *continuous random variables. Then if $g, h : \mathbb{R} \to \mathbb{R}$ we have*

$$\mathbb{E}\big[g(X)h(Y)\big] = \mathbb{E}\big[g(X)\big]\mathbb{E}\big[h(Y)\big]$$

22 Bivariate Distributions of the Continuous Type

22.1 Bivariate Distributions of the Continuous Type

Definition 22.1.

- Let X, Y be continuous random variables.
- We define the **covariance** of X and Y to be

$$cov(X, Y) = \mathbb{E}\big[\big(X - \mathbb{E}[X]\big)\big(Y - \mathbb{E}[Y]\big)\big] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

- We define the **correlation coefficient** of X and Y to be

$$\rho(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)\text{var}(Y)}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

- We have $-1 \le \rho(X, Y) \le 1$

Definition 22.2.

- Let X, Y be continuous random variables with joint PDF $f_{X,Y}(x, y)$
- Given $x \in \mathbb{R}$ so that $f_X(x) > 0$, we define the continuous random variable $Y|x$ with PDF

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x, y)}{f_X(x)}$$

Theorem 22.1. *If X, Y are continuous random variables and $x \in \mathbb{R}$ so that $f_X(x) > 0$ then*

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

Definition 22.3.

- Let X, Y be continuous random variables.
- Let $g(x) = \mathbb{E}[Y|x]$
- We define the **conditional expectation** to be the **random variable**

$$\mathbb{E}[Y|X] = g(X)$$

- We have the **Law of Iterated Expectation**:

$$\mathbb{E}\big[\mathbb{E}[Y|X]\big] = \mathbb{E}[Y]$$

Definition 22.4.

- Let X, Y be continuous random variables.
- Let $h(x) = \mathrm{var}(Y|x)$.
- We define the **conditional variance** to be the **random variable**

$$\mathrm{var}(Y|X) = h(X)$$

- We have the **Law of Total Variance**:

$$\mathbb{E}\big[\mathrm{var}(Y|X)\big] + \mathrm{var}\big((\mathbb{E}[Y|X]\big) = \mathrm{var}(Y)$$

23 Functions of Random Variables

23.1 Functions of Random Variables

Theorem 23.1. *Let X be a discrete random variable taking values in a set $S \subseteq \mathbb{R}$ with PMF $p_X(x)$ and let $u : S \to \mathbb{R}$. The PMF of $Y = u(X)$ is then*

$$p_Y(y) = \sum_{x \in \{\, x \in S : u(x) = y \,\}} p_X(x)$$

Theorem 23.2. *Let X be a continuous random variable with PDF $f_X(x)$. Let $S \subseteq \mathbb{R}$ so that $f_X(x) = 0$ for all $x \in \mathbb{R} \setminus S$. Let $u : \mathbb{R} \to \mathbb{R}$ be a smooth function satisfying $u'(x) > 0$ or $u'(x) < 0$ for all $x \in S$. Then $Y = u(X)$ has PDF*

$$f_Y(y) = \left| \frac{d}{dy} u^{-1}(y) \right| f_X(u^{-1}(y))$$

Theorem 23.3. *Let X, Y be discrete random variables taking values in sets $S_X, S_Y \subseteq \mathbb{R}$ and let $S = S_X \times S_Y$. Let $u : \mathbb{R}^2 \to \mathbb{R}^2$ and $(Z, W) = u(X, Y)$. If X, Y have joint PMF $p_{X,Y}(x, y)$ then Z, W have joint PMF*

$$p_{Z,W}(z, w) = \sum_{(x,y) \in \big\{\, (x,y) \in S : u(x,y) = (z,w) \,\big\}} p_{X,Y}(x, y)$$

Theorem 23.4. *Let X, Y be continuous random variables. Let $u : \mathbb{R}^2 \to \mathbb{R}^2$ be smooth and invertible, with inverse $v(z, w)$. If X, Y have joint PDF $f_{X,Y}(x, y)$ then $(Z, W) = u(X, Y)$ have joint PDF*

$$f_{Z,W}(z, w) = f_{X,Y}(v(z, w)) \left| \frac{\partial(x, y)}{\partial(z, w)} \right|$$

24 Several Independent Random Variables

24.1 Several Independent Random Variables

Remark (Special Case (Repeated Trials of Some Experiment)). $X_1, X_2, \ldots, X_n$ are **independent and identically distributed** (i.i.d)

Definition 24.1.

- Let $X_1, X_2, \ldots, X_n$ be **discrete** random variables taking values in sets $S_1, S_2, \ldots, S_n \subseteq \mathbb{R}$ and let $S = S_1 \times S_2 \times \cdots \times S_n \subseteq \mathbb{R}^n$.
- We may define their **joint PMF** to be

$$p_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = \mathbb{P}[X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n]$$

- We may define the **marginal PMF** of X_j to be

$$\begin{aligned} p_{X_j}(x_j) &= \mathbb{P}[X_j = x_j] \\ &= \sum_{x_1 \in S_1} \cdots \sum_{x_{j-1} \in S_{j-1}} \sum_{x_{j+1} \in S_{j+1}} \cdots \sum_{x_n \in S_n} p_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) \end{aligned}$$

- We say that $X_1, X_2, \ldots, X_n$ are **independent** if

$$p_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = p_{X_1}(x_1)\, p_{X_2}(x_2) \cdots p_{X_n}(x_n)$$

  for all $(x_1, x_2, \ldots, x_n)$

Definition 24.2.

- Let $X_1, X_2, \ldots, X_n$ be **continuous** random variables.
- We define their **joint PMF**

$$f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) \geq 0$$

  so that given any $A \subseteq \mathbb{R}^n$ we have

$$\mathbb{P}[(X_1 \ldots, X_n) \in A] = \int_A f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)\, dx_1 \ldots dx_n$$

- We define the **marginal PDF** of X_j by

$$f_{X_j}(x_j) = \int_{\mathbb{R}^{n-1}} f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)\, dx_1 \ldots dx_{j-1} dx_{j+1} \ldots dx_n$$

- We say that $X_1, X_2, \ldots, X_n$ are **independent** if

$$f_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = f_{X_1}(x_1)\, f_{X_2}(x_2) \cdots f_{X_n}(x_n)$$

  for all $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$

Definition 24.3.

- Let $u : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$
- If $X_1, \ldots, X_n$ are **discrete** random variables we define

$$\mathbb{E}\big[u(X_1, \ldots, X_n)\big] = \sum_{(x_1,\ldots,x_n) \in S} u(x_1, \ldots, x_n)\, p_{X_1,\ldots,X_n}(x_1, x_2, \ldots, x_n)$$

- If $X_1, \dots, X_n$ are **discrete** random variables we have

$$\mathbb{E}[g(X_j)] = \sum_{x_j \in S_j} g(x_j)\, p_{X_j}(x_j)$$

Definition 24.4.

- Let $u : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R} \to \mathbb{R}$
- If $X_1, \dots, X_n$ are **continuous** random variables we define

$$\mathbb{E}\big[u(X_1, \dots, X_n)\big] = \int_{\mathbb{R}^n} u(x_1, \dots, x_n)\, f_{X_1,\dots,X_n}(x_1, x_2, \dots, x_n)\, dx_1 \dots dx_n$$

- If $X_1, \dots, X_n$ are **continuous** random variables we have

$$\mathbb{E}[g(X_j)] = \int_{-\infty}^{\infty} g(x_j)\, f_{X_j}(x_j)\, dx_j$$

Theorem 24.1. *Let $X_1, \dots, X_n$ be discrete or continuous random variables.*

- *If $a \in \mathbb{R}$ then $\mathbb{E}[a] = a$.*
- *If $u, v : \mathbb{R}^n \to \mathbb{R}$ and $a, b \in \mathbb{R}$ then*

$$\begin{aligned} &\mathbb{E}\big[a\, u(X_1, \dots, X_n) + b\, v(X_1, \dots, X_n)\big] \\ &\qquad = a\, \mathbb{E}\big[u(X_1, \dots, X_n)\big] + b\, \mathbb{E}\big[v(X_1, \dots, X_n)\big] \end{aligned}$$

- *If $u(x_1, \dots, x_n) \leq v(x_1, \dots, x_n)$ for all $(x_1, \dots, x_n$ then*

$$\mathbb{E}\big[u(X_1, \dots, X_n)\big] \leq \mathbb{E}\big[v(X_1, \dots, X_n)\big]$$

Theorem 24.2. *Let $X_1, X_2, \dots, X_n$ be discrete or continuous random variables. Let $a_1, a_2, \dots, a_n \in \mathbb{R}$ and let*

$$Y = a_1 X_1 + \dots + a_n X_n$$

Then

$$\mathbb{E}[Y] = a_1 \mathbb{E}[X_1] + \dots + a_n \mathbb{E}[X_n]$$

Theorem 24.3. *Let $X_1, \dots, X_n$ be **independent** discrete or continuous random variables. Let $g_1, \dots, g_n : \mathbb{R} \to \mathbb{R}$. Then,*

$$\mathbb{E}\big[g_1(X_1) \dots g_n(X_n)\big] = \mathbb{E}[g_1(X_1)] \cdots \mathbb{E}[g_n(X_n)]$$

Theorem 24.4. *Let $X_1, X_2, \dots, X_n$ be **independent** discrete or continuous random variables. Let $a_1, a_2, \dots, a_n \in \mathbb{R}$ and let*

$$Y = a_1 X_1 + \dots + a_n X_n$$

Then

$$var(Y) = a_1^2 var(X_1) + \dots + a_n^2 var(X_n)$$

Definition 24.5.

- Let $X_1, X_2, \dots, X_n$ are **independent and identically distributed**

- Define the sample sum
$$S_n = X_1 + \cdots + X_n$$

- Define the sample average
$$\overline{X} = \frac{1}{n}(X_1 + \cdots + X_n)$$

25 Chebyshev's Inequality and Convergence in Probability

25.1 Chebyshev's Inequality and Convergence in Probability

Definition 25.1 (Convergence of Real Numbers)**.** Given a sequence of real numbers $(x_n)_{n\geq 1} \subseteq \mathbb{R}$ and a real number $x \in \mathbb{R}$ we say that
$$x_n \to x \quad \text{as} \quad n \to \infty$$
if, given any $\varepsilon > 0$ there exists some $N \geq 1$ so that for all $n \geq N$ we have
$$|x_n - x| < \varepsilon$$

Definition 25.2 (Convergence of Random Variables)**.**

- Let $(X_n)_{n\geq 1}$ be a sequence of random variables and X be another random variable
- There are **several types of convergence** for random variables
- We say that
$$X_n \to X \quad \textbf{in probability} \text{ as} \quad n \to \infty$$
if, given any $\varepsilon > 0$ we have
$$\mathbb{P}\big[|X_n - X| \geq \varepsilon\big] \to 0 \quad \text{as} \quad n \to \infty$$

Theorem 25.1 (The Weak Law of Large Numbers)**.** *Let $X_1, X_2, \ldots$ be an i.i.d. sequence of random variables with $\mathbb{E}[|X|] < \infty$. Then,*
$$\overline{X} = \frac{1}{n}\sum_{j=1}^{n} X_j \to \mu \quad \textit{in probability as} \quad n \to \infty$$

Theorem 25.2 (Markov's Inequality)**.** *Let $X \geq 0$ be a* ***non-negative*** *random variable. Then, given $a > 0$ we have*
$$\mathbb{P}[X \geq a] \leq \frac{\mathbb{E}[X]}{a}$$

Theorem 25.3 (Markov's Inequality v.2)**.** *Let $X \geq 0$ be a* ***non-negative*** *random variable. Then, given $a > 0$ and an integer $k \geq 1$ we have*
$$\mathbb{P}[X \geq a] \leq \frac{\mathbb{E}[X^k]}{a^k}$$

Theorem 25.4 (Chebyshev's Inequality)**.** *Let X be a random variable with mean μ and variance σ^2. Then, given $a > 0$ we have*
$$\mathbb{P}\big[|X - \mu| \geq a\big] \leq \frac{\sigma^2}{a^2}$$

Theorem 25.5 (The Chernoff Bound)**.** *Let X be a random variable. Then, given $a > 0$ we have*
$$\mathbb{P}[X \geq a] \leq \inf_{t>0}\big(e^{-ta}M_x(t)\big)$$

Theorem 25.6 (The Weak Law of Large Numbers)**.** *Let $X_1, X_2, \dots$ be an i.i.d. sequence of random variables with $\mathbb{E}[|X|] < \infty$. Then,*

$$\overline{X} = \frac{1}{n}\sum_{j=1}^{n} X_j \to \mu \quad \textit{in probability as} \quad n \to \infty$$

26 The Central Limit Theorem

26.1 The MGF Technique

Theorem 26.1. *Let $X_1, X_2, \dots, X_n$ be a sequence of* ***independent*** *random variables and $a_1, a_2, \dots, a_n \in \mathbb{R}$. Let*

$$Y = \sum_{j=1}^{n} a_j X_j$$

Then Y has MGF

$$M_Y(t) = \prod_{j=1}^{n} M_{X_j}(a_j t)$$

whenever it is well-defined.

Theorem 26.2. *Let X, Y be random variables with MGFs $M_X(t)$ and $M_Y(t)$. Suppose that for some $h > 0$ and all $t \in (-h, h)$ we have*

$$M_X(t) = M_Y(t)$$

Then X and Y are identically distributed.

Theorem 26.3. *Let $X_1, X_2, \dots, X_n$ be i.i.d. random variables with common MGF $M(t)$. Then*

$$M_{S_n}(t) = [M(t)]^n$$
$$M_{\overline{X}}(t) = [M(\tfrac{t}{n})]^n$$

26.2 Limiting MGFs

Definition 26.1.

- Let $X_1, X_2, \dots$ be a sequence of random variables.
- We say that

  $$X_n \to X \quad \textbf{in distribution} \text{ as } \quad n \to \infty$$

  if the CDF

  $$F_{X_n}(x) \to F_X(x) \quad \text{as} \quad n \to \infty$$

  for all $x \in \mathbb{R}$ so that $F_X(x)$ is continuous at x.

Theorem 26.4. *Let $X_1, X_2, \dots$ and X be random variables. Suppose that for some $h > 0$ and all $t \in (-h, h)$ we have*

$$M_{X_n}(t) \to M_X(t) \quad \textit{as} \quad n \to \infty$$

Then $X_n \to X$ ***in distribution*** *as $n \to \infty$*

26.3 The Central Limit Theorem

Theorem 26.5 (The Central Limit Theorem)**.** *Let $X_1, X_2, \ldots$ be an i.i.d. sequence of random variables with finite mean μ and variance σ^2. Then,*

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \to \mathcal{N}(0,1) \quad \textit{in distribution as} \quad n \to \infty$$

Remark.

- The CLT says that
$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \to \mathcal{N}(0,1) \quad \text{in distribution as} \quad n \to \infty$$
- Recall that if $Z \sim \mathcal{N}(0,1)$ then $\sigma\, Z + \mu \sim \mathcal{N}(\mu, \sigma^2)$
- As a consequence, the CLT says that
$$\overline{X} \approx \mathcal{N}(\mu, \tfrac{\sigma^2}{n})$$
when n is sufficiently large.
- Note that $\mathbb{E}[\overline{X}] = \mu$ and $\mathrm{var}(\overline{X}) = \frac{\sigma^2}{n}$
- Similarly $S_n \approx \mathcal{N}(n\mu, n\sigma^2)$