# MATH630: Probability Theory

Jonathan Ma
johnma@udel.edu

December 21, 2024

## Contents

# 1 Week 1

## 1.1 Lecture 1. Wed Aug 28

### 1.1.1 The Axioms of Probability

**Definition 1.1** (Sample space)**.** The sample space of a random experiment is the set of all outcomes $\Omega$.

**Definition 1.2** (Event)**.** Given a sample space $\Omega$, an event is a subset of $\Omega$. We think of collections of events as collections of sets, denoted by $\mathcal{F}$. We call $\mathcal{F}$ an $\sigma$-algebra.

**Definition 1.3** (Sigma algebra)**.** We say that a collection of sets $\mathcal{F}$ is a $\sigma$-algebra if the following conditions are satisfied:

1. $\emptyset \in \mathcal{F}$,
2. if $A \in \mathcal{F}$, then $A^C \in \mathcal{F}$,
3. if $A_1, \ldots, A_n$ is a sequence of sets in $\mathcal{F}$, then $\bigcup_{i=1}^{n} A_i \in \mathcal{F}$. We may take $n$ to be infinite.

**Proposition 1.4** (De Morgan's Laws)**.** *Let $\{A_i\}$ be some collection of sets. Then $(\bigcup_i A_i)^C = \bigcap_i A_i^C$.*

*Proof.* Let $x \in (\bigcup_i A_i)^C$. Therefore, $x \notin \bigcup_i A_i$. Then $x$ cannot be an element of any $A_i$. Therefore, $x$ must be an element of each $A_i^C$. This implies that $x \in \bigcap_i A_i^C$. Now let $y \in \bigcap_i A_i^C$. If $y$ is not a member of any $A_i$, then it must be in the complement of the union of each $A_i$. □

**Definition 1.5** (Probability)**.** For an event $A$, the probability of $A$ occuring is $P(A) = \frac{|A|}{|\Omega|}$.

**Theorem 1.6** (Axioms of probability)**.** *Let $\Omega$ be a nonempty sample space, and let $\mathcal{F}$ be a $\sigma$-algebra of sets of $\Omega$. We say $P\colon \mathcal{F} \to [0,1]$ is a probability measure if the following conditions are satisfied:*

1. *$P(\Omega) = 1$,*
2. *if $\{A_i\} \in \mathcal{F}$ are mutually exclusive (disjoint) sets, then $P(\bigcup_i A_i) = \sum_i P(A_i)$.*

# 2 Week 2

## 2.1 Lecture 2. Wed Sep 4

**Theorem 2.1.** *If $A, B \in \mathcal{F}$,*

1. *$P(\emptyset) = 0$.*
2. *If $A \subset B, P(A) \leq P(B)$.*
3. *The probability $P(A^C) = 1 - P(A)$.*
4. *The probability $P(A - B) = P(A \cap B^C) = P(A) - P(A \cap B)$.*
5. *The probability $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.*

*Proof.* We will prove some of these results, remaining are homework assignments.

1. Since $1 = P(\Omega) = P(\Omega \cup \emptyset) = P(\Omega) + P(\emptyset) \Rightarrow P(\emptyset) = 0$.

3. We know $\Omega = A \cup A^C$, and that $A, A^C$ are disjoint, so that $P(\Omega) = P(A) + P(A^C) \Rightarrow P(A) = 1 - P(A^C)$.

5. We have that $P(A \cup B) = P((A \cup B) \cap \Omega) = P((A \cup B) \cap (A \cup A^C)) = P((A \cup B) \cap A) \cup (A \cup B) \cap A^C)) = P(A \cup (B \cap A^C)) = P(A) + P(B \cap A^C) = P(A) + P(B) - P(A \cap B)$.

□

**Theorem 2.2** (Principle of inclusion exclusion)**.** *The probability of a finite union of events*

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) + \ldots + (-1)^{n+1} P(A_1 \cap A_2 \cap \ldots \cap A_n)$$

.

**Theorem 2.3.** *If $\Omega$ is finite, and each outcome is equally likely, then $P(A) = \frac{|A|}{|\Omega|}$, for any event $A$.*

*Proof.* Let $\Omega = \{\omega_1, \ldots, \omega_n\}$. Then $1 = P(\Omega) = P(\bigcup_{i=1}^{n} \{\omega_i\}) = \sum_{i=1}^{n} P(\{\omega_i\}) = nP(\{\omega_i\}) \Rightarrow P(\{\omega_i\}) = \frac{1}{n} = \frac{1}{|\Omega|}$. Then, $P(A) = \sum_{i:\omega_i \in A} P(\{\omega_i\}) = \frac{|A|}{|\Omega|}$ □

# 3 Week 3

## 3.1 Lecture 3. Mon Sep 9

### 3.1.1 Conditional Probability

**Definition 3.1.** Let $A, B \in \mathcal{F}$. The conditional probability that $A$ occurs given that $B$ occurs is defined by $P(A \mid B) = \dfrac{P(A \cap B)}{P(B)}$, $P(B) > 0$.

**Theorem 3.2.** *The function $P(\cdot \mid B)$ is a probability measure.*

**Exercise 3.1.** A fair die is rolled twice. Find the probability that one of the numbers is a $4$ given that the sum of the two dice is $7$.

### 3.1.2 Law of Total Probability, Bayes' Theorem

**Theorem 3.3** (Law of total probability)**.** *Let $\{B_i\}_{i=1}^{\infty}$ be a countable sequence of disjoint events such that $\Omega = \bigcup_{i=1}^{\infty} B_i$. Let $A$ be any event, then*

$$P(A) = \sum_{i=1}^{\infty} P(A \mid B_i) P(B_i).$$

*Proof.* Notice that $A = \bigcup_{i=1}^{\infty} (A \cap B_i)$. Then, since each $B_i$ is disjoint, $P(A) = \sum_{i=1}^{\infty} P(A \cap B_i) = \sum_{i=1}^{\infty} P(A \mid B) P(B)$. □

**Corollary 3.3.1.** *Let $A, B \in \mathcal{F}$. Then $P(A) = P(A \mid B)P(B) + P(A \mid B^C)P(B^C)$.*

**Theorem 3.4** (Bayes' theorem)**.** *Let $\{B_i\}_{i=1}^{\infty}$ be a disjoint countable set of events, with $\Omega = \bigcup_{i=1}^{\infty} B_i$. Then for any event $A$,*

$$P(B_j \mid A) = \frac{P(A \mid B_j) P(B_j)}{\sum_{i=1}^{\infty} P(A \mid B_i) P(B_i)}.$$

## 3.2 Lecture 4. Wed Sep 11

**Example 3.5.** There are two identical urns. Urn I contains 3 red balls, and 2 black balls. Urn II contains 5 red balls, and 4 black balls. What is the probability that a randomly chosen ball from one of the urns is red.

### 3.2.1 Random Variables

**Definition 3.6** (Random variable)**.** Let $(\Omega, \mathcal{F}, P)$ be a probability space. A random variable $X$ is a real valued function $X : \Omega \to \mathbb{R}$ where $X^{-1}((-\infty, a]) = \{\omega \in \Omega \mid X(\omega) \leq a\} \in \mathcal{F}$, for $a \in \mathbb{R}$.

**Definition 3.7.** The distribution function $F$ of $X$ is defined by $F(x) = P\{\omega \mid X(\omega) \leq x\}$. We write this as $F(x) = P(X \leq x)$.

**Proposition 3.8.** *For all $x, y \in \mathbb{R}$,*

1. $0 \leq F(x) \leq 1$,
2. $x \leq y \Rightarrow F(x) \leq F(y)$,
3. $\lim_{x \to -\infty} F(x) = 0$, $\lim_{x \to \infty} F(x) = 1$,
4. $\lim_{y \to x^+} F(y) = F(x)$,
5. $F(x-) := \lim_{y \to x^-} F(y)$ *exists,*
6. $F$ *has at most countably many discontinuities.*

**Lemma 3.9.** *Let $A_1 \subset A_2 \subset \ldots \subset \ldots$ be an sequence of increasing sets in $\mathcal{F}$. Then*

$$\lim_{n \to \infty} P(A_n) = P\Big( \bigcup_{n=1}^{\infty} A_n \Big).$$

*If $B_1 \supset B_2 \supset \ldots \supset \ldots$ is a sequence of decreasing sets in $\mathcal{F}$, then*

$$\lim_{n \to \infty} P(B_n) = P\Big( \bigcap_{n=1}^{\infty} B_n \Big).$$

# 4 Week 4

## 4.1 Lecture 5. Mon Sep 16

*Lemma 3.9*. Notice that $\bigcup_{i=1}^{\infty} A_i = A_1 \cup (A_2 - A_1) \cup (A_3 - A_2) \cup \ldots$. All of the sets on the RHS are disjoint, so that

$$\begin{aligned}
P\left(\bigcup_{i=1}^{\infty} A_i\right) &= P(A_1) + \sum_{i=2}^{\infty} P(A_i - A_{i-1}) \\
&= P(A_1) + \sum_{i=2}^{\infty} P(A_i) - P(A_i \cap A_{i-1}) \\
&= P(A_1) + \sum_{i=2}^{\infty} P(A_i) - P(A_{i-1}) \\
&= P(A_1) + \lim_{N\to\infty} \sum_{i=2}^{N} \Big[P(A_i) - P(A_{i-1})\Big] \\
&= P(A_1) + \lim_{N\to\infty} \Big[P(A_2) - P(A_1) + P(A_3) - P(A_2) + \ldots + P(A_N) - P(A_{N-1})\Big] \\
&= \lim_{N\to\infty} P(A_N).
\end{aligned}$$

□

*Proposition 3.8*.
(ii). Suppose $x \le y$. Define $B_x = \{\omega \mid X(\omega) \le x\}$ and $B_y = \{\omega \mid X(\omega) \le y\}$. Notice that $B_x \subset B_y$, so that $P(B_x) \le P(B_y) \Rightarrow P(\{\omega \mid X(\omega) \le x\}) \le P(\{\omega \mid X(\omega) \le y\}) \Rightarrow F(x) \le F(y)$.
(iii). Define $B_n := \{\omega \mid X(\omega) \le n\}$, for all $n \ge 1$. Therefore, $\bigcup_{n=1}^{\infty} B_n = \Omega$, so that

$$\begin{aligned}
P\left(\bigcup_{n=1}^{\infty} B_n\right) = P(\Omega) = 1 \\
\Rightarrow \lim_{n\to\infty} P(B_n) = 1 \\
\lim_{n\to\infty} F(n) = 1.
\end{aligned}$$

(iv). Let $\{y_n\}$ be an decreasing sequence such that $\lim_{n\to\infty} y_n = x$. Set $B_{y_n} := \{\omega \mid X(\omega) \le y_n\}$. Note

that $B_{y_1} \supset B_{y_2} \supset \ldots$, so that $\bigcap_{i=1}^{\infty} B_{y_n} = B_x$. Then

$$\begin{aligned} P\left(\bigcap_{i=1}^{\infty} B_{y_i}\right) &= P(B_x) \\ \lim_{n\to\infty} P(B_{y_n}) &= P(B_x) \\ \lim_{n\to\infty} F(y_n) &= F(x) \\ \Rightarrow \lim_{y\to x^+} F(y) &= F(x). \end{aligned}$$

□

**Theorem 4.1.** *The following hold true:*

1. $P(X < x) = F(x^-)$,
2. $P(X = x) = F(x) - F(x^-)$,
3. $P(a \le x \le b) = F(b) - F(a^-)$,
4. $P(X > x) = 1 - F(x)$.

*Proof.* (1). Set $B_{x-\frac{1}{n}} := \{\omega \mid X(\omega) \le x - \frac{1}{n}\}$. Notice that $\{\omega \mid X(\omega) < x\} = \bigcup_{n=1}^{\infty} B_{x-\frac{1}{n}}$. Also, $\{B_{x-\frac{1}{n}}\}$ is an increasing sequence, so that

$$\begin{aligned} P\Big\{\omega \mid X(\omega < x)\Big\} &= P\left(\bigcup_{n=1}^{\infty} B_{x-\frac{1}{n}}\right) \\ &= \lim_{n\to\infty} P(B_{x-\frac{1}{n}}) \\ &= \lim_{n\to\infty} F\left(x - \frac{1}{n}\right) \\ &= F(x^-). \end{aligned}$$

□

## 4.2 Lecture 6. Wed Sep 18

### 4.2.1 Distributions and Independence

**Definition 4.2.** Let $I_A(x) = \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases}$. Call $I_A(x)$ an indicator function.

**Definition 4.3.** Let $\Omega = [0, 1]$, and $X(\omega) = \omega \in [0, 1]$. Let $P$ be a measure that gives the length of a set $A$ : $P[a \le X \le b] = P(x \in [a, b]) = b - a$. Then $X$ has a uniform distribution over $[0, 1]$.

**Definition 4.4.** Two events $A, B \in \mathcal{F}$ are said to be independent if $P(A \cap B) = P(A)P(B)$.

**Theorem 4.5.** *If $A, B$ are independent, and $P(B) > 0$, then $P(A \mid B) = P(A)$.*

*Proof.* We know $P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A)$. □

**Example 4.6.** If $A, B$ are disjoint are the independent? No, since $P(A \cap B) = P(A)P(B) \Rightarrow P(A) = 0$ or $P(B) = 0$.

**Example 4.7.** Is it possible for $A$ to be independent of itself? Now, $P(A \cap A) = P(A)P(A) \Rightarrow P(A) = P(A)^2 \Rightarrow P(A) = 0 \vee P(A) = 1$.

**Example 4.8.** If $A, B$ are independent, prove that $A, B^C$ are also independent.

*Proof.* Now, $P(A \cap B^C) = P(A) - P(A \cap B) = P(A) - P(A)P(B) = P(A)(1 - P(B)) = P(A)P(B^C)$. □

**Definition 4.9.** A sequence of events $\{A_n\}$ is said to be independent if for any finite subsequence $i_1, i_2, \ldots, i_k$,

$$P(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}) = \prod_{n=1}^{k} P(A_{i_n}).$$

**Definition 4.10.** The random variables $X, Y$ are independent random variables if for all $x, y$, $P(X \leq x, Y \leq y) = P(X \leq y)P(Y \leq y)$.

**Theorem 4.11.** *If $X, Y$ are independent, then*

$$P(X \in A, Y \in A) = P(X \in A)P(Y \in A),$$

*for sets of the form $A = (p, q\,], B = (r, s\,]$.*

*Proof.* Let $A = (-\infty, q\,], B = (r, s\,]$. Then

$$\begin{aligned}
P\Big[x \in (-\infty, q\,], Y \in (r, s\,]\Big] &= P\Big[x \leq q, r < Y \leq s\Big] \\
&= P\Big[x \leq q, Y \leq s\Big] - P\Big[X \leq q, Y \leq r\Big] \\
&= P\Big[X \leq q\Big] P\Big[Y \leq s\Big] - P\Big[X \leq q\Big] P\Big[Y \leq r\Big] \\
&= P\Big[x \leq q\Big] \Big(P\Big[Y \leq s\Big] - P\Big[Y \leq r\Big]\Big) \\
&= P(x \leq q) P(r < Y \leq s).
\end{aligned}$$

□

**Definition 4.12.** Let $\{X_n\}$ be a sequence of random variables. Then $\{X_n\}$ is independent if for any finite subsequence $i_1, \ldots, i_k$ and $x_1, \ldots, x_n \in \mathbb{R}$,

$$P\Big[X_{i_1} \leq x_i, \ldots, X_{i_k} \leq x_{i_k}\Big] = \prod_{n-1}^{k} P\Big[X_{i_n} \leq x_{i_n}\Big].$$

**Definition 4.13.** We say $X$ is a discrete random variable if there exists a countable set $\{x_n\}$ such that $\sum_n P(X = x_n) = 1$. Or, $P\Big(X \notin \bigcup_{i=1}^{\infty} \{x_n\}\Big) = 0$.

**Definition 4.14.** We say $X$ is a continuous random variable if there exists a function $f : \mathbb{R} \to \mathbb{R}$ such that $F(x) = P(X \leq x) = \int_{-\infty}^{\infty} f(x)\,dx$.

**Note.** The function $f$ is called the density or probability density function (pdf) of $X$.

**Note.** There are other types of distributions, for example, mixed distributions.

**Definition 4.15.** The probability mass function of a *discrete* random variable $X$ is given by $p(x) = P(X = x)$, for all $x \in \mathbb{R}$.

**Note.** We must have $p(x) = 0$ except at countable many points. Also, $\sum_{x \,|\, p(x)>0} p(x) = 1$, and $p(x) \geq 0, \forall x \in \mathbb{R}$.

**Example 4.16.** Consider flipping a fair coin twice. Define $X$ to be the number of heads observed minus the number of tails observed. Therefore, $X \in \{-2, 0, 2\}$. Then

$$p(x) = \begin{cases} 0 & x \notin \{-2, 0, 2\} \\ \frac{1}{4} & x = -2 \\ \frac{1}{2} & x = 0 \\ \frac{1}{4} & x = 2. \end{cases}$$

The distribution function of $X$ is written

$$F(x) = \begin{cases} 0 & x < -2 \\ \frac{1}{4} & x < 0 \\ \frac{3}{4} & x < 2 \\ 1 & x \geq 2. \end{cases}$$

**Example 4.17.** Is the following function a valid probability mass function:

$$p(x) = \begin{cases} \log\left(\frac{x+1}{x}\right) & 1 \leq x \leq 9, x \in \mathbb{Z} \\ 0 & \text{otherwise.} \end{cases}$$

**Proposition 4.18.** *There exists an $f$ such that the distribution of a random variable $X$ is $F(x) = \int_{-\infty}^{x} f(y)\,dy$. This function $f$ is also at least 0, for all $x \in \mathbb{R}$, and $f$ is monotonically increasing. Whenever the derivative of $F$ exists, we also have $F'(x) = f(x) \geq 0$.*

**Proposition 4.19.** *We have* $\int_{-\infty}^{\infty} f(x)\,dx = \lim_{x\to\infty} F(x) = 1$.

# 5 Week 5

## 5.1 Lecture 7. Mon Sep 23

### 5.1.1 Expected Value & Variance

**Definition 5.1.** The expected value of a discrete random variable $X$, denoted by $\mathbf{E}[X]$, is given by

$$\mathbf{E}[X] = \sum_{x|p(x)>0} xP(X = x),$$

when $\sum_{x|p(x>0)} |x|P(X = x) < \infty$.

**Example 5.2.** Roll a fair die twice. Let $X$ be the number of 1s observed. What is the expected value of $X$? Note that $P(X=0)=\frac{25}{36}, P(X=1)=\frac{10}{36}$, and $P(X=2)=\frac{1}{36}$. Then $\mathbf{E}[X]=0\cdot\frac{25}{36}+1\cdot\frac{10}{36}+2\cdot\frac{1}{36}=\frac{1}{3}$.

**Example 5.3.** Suppose the pmf of $X$ is given by

$$p(n)=\frac{1}{n(n+1)},\ n=1,2,3,\ldots.$$

Find $\mathbf{E}[X]$. Note that

$$\mathbf{E}[X]=\sum_{n=1}^{\infty} n\cdot p(n)=\sum_{n=1}^{\infty} n\frac{1}{n(n+1)}\to\infty,$$

$\therefore \mathbf{E}[X]$ does not exist.

**Theorem 5.4.** *If $X$ is a discrete random variable and $g:\mathbb{R}\to\mathbb{R}$, then*

$$\mathbf{E}[g(X)]=\sum_{x|p(x)>0} g(x)P(X=x),$$

*when $\sum|g(x)|p(x)<\infty$.*

*Proof.* Let $\{g_j\}$ denote all the values $g(X)$ takes. Denote $Y=g(X)$, and $A_j=\{x\mid g(x)=y_j\}$. Then $g(A_j)=\{y_j\}$. Now,

$$\begin{aligned}\mathbf{E}[Y]&=\sum_j y_j P(Y=y_j)\\&=\sum_j y_j\sum_{x\in A_j} P(X=x)\\&=\sum_j\sum_{x\in A_j} g(x)P(X=x)\\&=\sum_x g(x)P(X=x). && (A_i\cap A_j=\varnothing, i\neq j)\end{aligned}$$

□

**Theorem 5.5.** *The following hold true.*

1. *Expectation is linear:* $\mathbf{E}[X+Y]=\mathbf{E}[X]+\mathbf{E}[Y]$.
2. *Expectation is homogeneous:* $\mathbf{E}[aX]=a\,\mathbf{E}[X]$.
3. *For* $x\le y, \mathbf{E}[X]\le\mathbf{E}[Y]$.
4. *For constant* $k\in\mathbb{R}, \mathbf{E}[k]=k$.
5. *For the indicator function* $I_A$, $\mathbf{E}[I_A]=P(A)$.

*Proof of linearity.* Let $\{x_i\}$ denote the values of $x$, and let $\{y_j\}$ denote the value of $Y$. Set $Z:=X+Y$. Then let $\{z_k\}$ be the values of $Z$. Set $A_k:=\{(i,j)\mid x_i+y_j=z_k\}$. Then we have

$$P(X+Y=z_k)=\sum_{(i,j)\in A_k} P(X=x_i,Y=y_j).$$

Now,

$$\begin{aligned}
\mathbf{E}[Z] &= \sum_k z_k P(X+Y=z_k) \\
&= \sum_k z_k \sum_{(i,j)\in A_k} P(X=x_i, Y=y_j) \\
&= \sum_k \sum_{(i,j)\in A_k} (x_i+y_j) P(X=x_i, Y=y_j) \\
&= \sum_i \sum_j (x_i+y_j) P(X=x_i, Y=y_j) \\
&= \sum_i \sum_j x_i P(X=x_i, Y=y_j) + y_j P(X=i, Y=y_j) \\
&= \left( \sum_i \sum_j x_i P(X=x_i, Y=y_j) \right) + \left( \sum_i \sum_j y_j P(X=i, Y=y_j) \right) \\
&= \left( \sum_i \sum_j x_i P(X=x_i \mid Y=y_j) P(y_j) \right) + \left( \sum_j \sum_i y_j P(Y=y_j \mid X=x_i) P(x_i) \right) \\
&= \left( \sum_i x_i P(X=x_i) \right) + \left( \sum_j y_j P(Y=y_j) \right).
\end{aligned}$$

□

**Definition 5.6.** The variance of a random variable $X$ is given by

$$\mathbf{Var}[X] = \mathbf{E}\left[(X - E(X))^2\right].$$

**Theorem 5.7.** *The variance of a random variable is also given by*

$$\mathbf{Var}[X] = \mathbf{E}\left[X^2\right] - (\mathbf{E}[X])^2.$$

**Exercise 5.1.** *What is the "best" prediction of a random variable $X$?* Find $t \in \mathbb{R}$ such that $\mathbf{E}\left[(X - t^2)\right]$ is minimized.

*Solution.* Call $R(t) := \mathbf{E}\left[(X-t)^2\right]$. Then

$$\begin{aligned}
R(t) &= \mathbf{E}\left[(x-t)^2\right] \\
&= \mathbf{E}\left[X^2 - 2tX + t^2\right] \\
&= \mathbf{E}\left[X^2\right] - 2t\,\mathbf{E}[X] + t^2 \\
R'(t) &= -2\,\mathbf{E}[X] + 2t = 0 \\
&\Rightarrow t = \mathbf{E}[X].
\end{aligned}$$

In statistics, the best predictor or estimator of $X$ is $\mathbf{E}[X]$. □

**Theorem 5.8.** *The following hold true:*

1. $\mathbf{Var}[aX+b] = a^2\,\mathbf{Var}[X]$,

2. $\mathbf{Var}[k] = 0, \forall\, k \in \mathbb{R}$.

**Note.** In general, $\mathbf{Var}[X+Y] \neq \mathbf{Var}[X] + \mathbf{Var}[Y]$.

*Proof of 1.* We have

$$\begin{aligned}\mathbf{Var}[aX+b] &= \mathbf{E}\left[(aX+b-\mathbf{E}[aX+b])^2\right] \\ &= \mathbf{E}\left[(aX+b-a\,\mathbf{E}[X])-b)^2\right] \\ &= a^2\,\mathbf{E}\left[(X-\mathbf{E}[X])^2\right] \\ &= a^2\,\mathbf{Var}[X].\end{aligned}$$

□

### 5.1.2 More Distributions

**Definition 5.9** (Bernoulli distribution)**.** We say $X$ is a Bernoulli random variable if $P(X=1)=p$, and $P(X=0)=1-p$. We write $X \sim \mathrm{Ber}(p)$.

**Theorem 5.10.** *If $X \sim \mathrm{Ber}(p)$, $\mathbf{E}[X]=p$, and $\mathbf{Var}[X]=p(1-p)$.*

**Definition 5.11** (Binomial distribution)**.** Consider $n$ independent Bernoulli random variables $X_1, X_2, \ldots, X_n$. Set

$$X = \sum_{i=1}^{n} X_i = \text{ number of successes of } n \text{ Bernoulli trials.}$$

Then $X$ is a Binomial random variable, written $X \sim \mathrm{Bin}(n,p)$.

**Example 5.12.** Suppose a fair coin is flipped 100 times. Then let $X$ be the number of heads observed. Then $X \sim \mathrm{Bin}(100, \frac{1}{2})$.

**Example 5.13.** Suppose we have $n$ independent trials, and we know $X$ is the random variable modeling the number of successes we observe. Then $P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$, for $k = 0, 1, \ldots, n$.

**Example 5.14.** We know that if $X \sim \mathrm{Bin}(n,p)$, then $X = X_1 + \ldots + X_n$, for $X_i \sim \mathrm{Ber}(p)$, $i \in [n]$. Then

$$\mathbf{E}[X] = \sum_{i=1}^{n} \mathbf{E}[X_i] = np.$$

**Theorem 5.15.** *If $X \sim \mathrm{Bin}(n,p)$, then $\mathbf{Var}[X] = np(1-p)$.*

## 5.2 Lecture 8. Wed Sep 25

**Definition 5.16** (Poisson distribution)**.** We say $X$ has a Poisson distribution if

$$P(X=k) = \frac{e^{-\lambda}\lambda^k}{k!}, \quad \lambda > 0, k = 0, 1, \ldots.$$

We write $X \sim \mathrm{Poisson}(\lambda)$.

**Theorem 5.17.** *If $X \sim \text{Poisson}(\lambda)$, then $\mathbf{E}[X] = \lambda$.*

*Proof.* We have

$$\begin{aligned}
\mathbf{E}[X] &= \sum xP(X = k) \\
&= \sum_{k=0}^{\infty} k \cdot \frac{e^{-\lambda}\lambda^k}{k!} \\
&= \sum_{k=1}^{\infty} \frac{e^{-1}\lambda^k}{(k-1)!} \\
&= e^{-\lambda}\lambda \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \\
&= e^{-\lambda}\lambda \underbrace{\left[1 + \lambda + \frac{\lambda^2}{2!} + \dots\right]}_{e^{\lambda}} \\
&= \lambda.
\end{aligned}$$

□

**Example 5.18.** Let $X \sim \text{Poisson}(\lambda), 0 < \lambda < 1$. Find $\mathbf{E}[X!]$. Find $\mathbf{E}\left[2^X\right]$.

*Proof.* We have

$$\begin{aligned}
\mathbf{E}[X!] &= \sum_{k=0}^{\infty} k! \frac{e^{-\lambda}\lambda^k}{k!} \\
&= e^{-\lambda} \sum_{k=0}^{\infty} \lambda^k \\
&= e^{-\lambda}\left[\frac{1}{1-\lambda}\right],
\end{aligned}$$

and

$$\begin{aligned}
\mathbf{E}\left[2^X\right] &= \sum_{k=0}^{\infty} \frac{2^k e^{-\lambda}\lambda^k}{k!} \\
&= e^{-\lambda} \sum_{k=0}^{\infty} \frac{(2\lambda)^k}{k!} \\
&= e^{\lambda}.
\end{aligned}$$

□

**Proposition 5.19.** *If $X \sim$ Poisson$(\lambda)$, then $\mathbf{Var}[X] = \lambda$.*

**Theorem 5.20** (Poisson approximation of binomial distribution)**.** *Let $X_n$ be a sequence of random variables, where each $X_n \sim B(n, \frac{\lambda}{n})$, where $\lambda > 0$. Then*

$$\lim_{n\to\infty} P(X_n = k) = \frac{e^{-\lambda}\lambda^k}{k!}, \forall k = 0, 1, 2, \dots.$$

*Proof.* We know

$$\begin{aligned} P(X_n = k) &= \binom{n}{k}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k} \\ &= \frac{n!}{k!(n-k)!}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n}\left(1-\frac{\lambda}{n}\right)^{-k} \\ &= \frac{n(n-1)(n-2)\cdots(n-k+1)}{n^k}\left(\frac{\lambda^k}{k!}\right)\left(1-\frac{\lambda}{n}\right)^{n}\left(1-\frac{\lambda}{n}\right)^{-k} \\ \lim_{n\to\infty} P(X_n = k) &= \frac{e^{-\lambda}\lambda^k}{k!}. \end{aligned}$$

□

**Note.** When $p$ is small, we can approximate the binomial distribution with the Poisson distribution.

**Definition 5.21** (Geometric distribution)**.** Consider a sequence of independent events $\{A_n\}$ each with the same probability of success. Then let $X$ be a random variable equal to the number of trials needed to observe the first success amongst $\{A_n\}$. We write $X \sim \mathrm{Geom}(p)$.

**Proposition 5.22.** *For $X \sim \mathrm{Geom}(p)$, $P(X = k) = (1-p)^{k-1}p$, for all $k = 1, 2, \ldots$. Also, $P(X \geq k) = P($First $k-1$ trials are failures$) = (1-p)^{k-1}$.*

**Theorem 5.23.** *Suppose $X \geq 0$, and $X$ is discrete. Then*

$$\mathbf{E}[X] = \sum_x P(X \geq x).$$

**Corollary 5.23.1.** *If $X \sim \mathrm{Geom}(p)$. Then $\mathbf{E}[X] = \frac{1}{p}$.*

*Proof.* We have by the previous theorem,

$$\mathbf{E}[X] = \sum_{k=0}^{\infty}(X \geq x) = \sum_{k=1}^{\infty}(1-p)^{k-1} = \frac{1}{1-(1-p)} = \frac{1}{p}.$$

□

# 6 Week 6

## 6.1 Lecture 9. Mon Sep 30

### 6.1.1 Joint Distributions

**Definition 6.1.** The joint distribution function $F : \mathbb{R}^2 \to [0,1]$ is defined by $F(x,y) = P(X \leq x, Y \leq y), \forall x, y \in \mathbb{R}$. The probability mass function $p : \mathbb{R}^2 \to [0,1]$ is given by $p(x,y) = P(X = x, Y = y)$.

**Proposition 6.2.** *Let $A_x = \{X = x\}, B_y = \{Y = y\}$. Then $p(x,y) = P(A_x \cap B_y)$. Note that $\{X = x\} = \bigcup_y(\{X = x\} \cap \{Y = y\})$, so that*

$$P[\{X = x\}] = \sum_y P(A_x \cap B_y) = \sum_y p(x,y).$$

*Let $p_X(x) = \sum_y p(x,y)$. Similarly, let $p_Y(y) = \sum_x p(x,y)$. Then $p_X$ and $p_Y$ are called marginal distribution functions.*

**Theorem 6.3.** *We have* $\mathbf{E}[g(X,Y)] = \sum_x \sum_y g(x,y)p(x,y)$.

**Theorem 6.4.** *The events* $x$ *and* $y$ *are independent if and only if* $p(x,y) = p_X(x) - p_Y(y), \forall x, y$.

**Theorem 6.5.** *If* $X, Y$ *are independent, then* $\mathbf{E}[XY] = \mathbf{E}[X]\,\mathbf{E}[Y]$.

*Proof.* Recall $\mathbf{E}[I_A] = P(A)$. Then

$$\begin{aligned} XY &= \sum_x \sum_y xy I_{A_x \cap I_{B_y}} \\ \mathbf{E}[XY] &= \sum_x \sum_y xy P(A_x \cap B_y) \\ &= \sum_x \sum_y p(x,y) \\ &= \sum_x \sum_y p_X(x) \cdot p_Y(y) \\ &= \sum_x x p_X(x) \Big( \sum_y y p_Y(y) \Big) \\ &= \mathbf{E}[X]\,\mathbf{E}[Y]\,. \end{aligned}$$

□

**Definition 6.6.** If $\mathbf{E}[XY] = \mathbf{E}[X]\,\mathbf{E}[Y]$, we say $X, Y$ are uncorrelated. Although this does not imply $X, Y$ are independent.

**Example 6.7.** Suppose $X \sim \mathrm{Geom}(\alpha), Y \sim \mathrm{Geom}(\beta)$, are independent. Define $Z = \min\{X, Y\}$. Find $P(Z = k)$.

*Proof.* We know

$$\begin{aligned} P(Z > k) &= P\{\omega \mid Z(\omega) > k\} \\ &= P\{\omega \mid \min\{X,Y\} > k\} \\ &= P\{\omega \mid X(\omega) > k, Y(\omega) > k\} \\ &= P\Big( \{\omega \mid X(\omega) > k\} \cap \{\omega \mid Y(\omega) > k\} \Big) \\ &= P(X > k) P(Y > k) \\ &= P(X \geq k+1) P(Y \geq k+1) \\ &= (1-\alpha)^k (1-\beta)^k. \end{aligned}$$

Now,

$$\begin{aligned} P(Z = k) &= P(Z > k-1) - P(Z - k) \\ &= [(1-\alpha)(1-\beta)]^{k-1} - [(1-\alpha)(1-\beta)]^k \\ &= [(1-\beta)(1-\beta)]^{k-1} [1 - (1-\alpha)(1-\beta)]. \end{aligned}$$

This implies $Z \sim \mathrm{Geom}(1 - (1-\alpha)(1-\beta))$. □

**Definition 6.8.** The covariance between $X$ and $Y$ is given by

$$\begin{aligned}\mathbf{Cov}\left[X,Y\right] &= \mathbf{E}\left[(X-\mathbf{E}\left[x\right])(Y-\mathbf{E}\left[Y\right])\right]\\ &= \mathbf{E}\left[XY\right]-\mathbf{E}\left[X\right]\mathbf{E}\left[Y\right].\end{aligned}$$

**Theorem 6.9.** *We have* $\mathbf{Var}\left[X+Y\right]=\mathbf{Var}\left[X\right]+\mathbf{Var}\left[Y\right]+2\,\mathbf{Cov}\left[X,Y\right]$.

*Proof.* We know

$$\begin{aligned}\mathbf{Var}\left[X\right] &= \mathbf{E}\left[X^2\right]-(\mathbf{E}\left[X\right])^2\\ \Rightarrow \mathbf{Var}\left[X+Y\right] &= \mathbf{E}\left[(X+Y)^2\right]-(\mathbf{E}\left[X+Y\right])^2\\ &= \mathbf{E}\left[X^2+2XY+Y^2\right]-[(\mathbf{E}\left[X\right])^2+2\,\mathbf{E}\left[X\right]\mathbf{E}\left[Y\right]+(\mathbf{E}\left[Y\right])^2]\\ &= \mathbf{E}\left[X^2\right]-(\mathbf{E}\left[X\right])^2+\mathbf{E}\left[Y^2\right]-(\mathbf{E}\left[Y\right])^2+2[\mathbf{E}\left[XY\right]-\mathbf{E}\left[X\right]\mathbf{E}\left[Y\right]]\\ &= \mathbf{Var}\left[X\right]+\mathbf{Var}\left[Y\right]+2\,\mathbf{Cov}\left[X,Y\right].\end{aligned}$$

□

**Proposition 6.10.** *If* $X,Y$ *are independent,* $\mathbf{Var}\left[X+Y\right]=\mathbf{Var}\left[X\right]+\mathbf{Var}\left[Y\right]$. *Also,* $\mathbf{E}\left[XY\right]=\mathbf{E}\left[X\right]\mathbf{E}\left[Y\right]$*, so* $\mathbf{Cov}\left[X,Y\right]=0$.

**Theorem 6.11.** *(CS for expectation) We have*

$$\mathbf{E}\left[(X,Y)\right]\leq\sqrt{\mathbf{E}\left[X^2\right]\mathbf{E}\left[Y^2\right]},$$

*if* $\mathbf{E}\left[X^2\right],\mathbf{E}\left[Y^2\right]<\infty$.

*Proof.* Let $t\in\mathbb{R}$, and $\mathbf{E}\left[X^2\right]\geq\mathbf{E}\left[Y^2\right]\geq 0$. Consider $\mathbf{E}\left[(tX+Y)^2\right]\geq 0$. Expanding, we find

$$\begin{aligned}\mathbf{E}\left[(tX+Y)^2\right]&\geq 0\\ \mathbf{E}\left[t^2X^2+2tXY+Y^2\right]&\geq 0\\ t^2\,\mathbf{E}\left[X^2\right]+2t\,\mathbf{E}\left[XY\right]+\mathbf{E}\left[Y^2\right]&\geq 0.\end{aligned}$$

Note that the quadratic above can only either have at most 1 real root. Therefore, its discriminant is less than or equal to $0$. Therefore,

$$\begin{aligned}(2\,\mathbf{E}\left[XY\right])^2-4(\mathbf{E}\left[X^2\right]\mathbf{E}\left[Y^2\right])&\leq 0\\ \Rightarrow \mathbf{E}\left[XY\right]^2&\leq\mathbf{E}\left[X^2\right]\mathbf{E}\left[Y^2\right].\end{aligned}$$

□

### 6.1.2 Sums of Random Variables

**Proposition 6.12.** *Let* $X,Y$ *be discrete. Consider* $Z=X+Y$*, and observe the set* $\{\omega\mid Z(\omega)=k\}=\{\omega\mid X(\omega)+Y(\omega)=k\}=\{X+Y=k\}$*. Then consider*

$$\begin{aligned}\{X+Y=k\}&=\bigcup_x^\infty\{\{X=x\}\cap\{Y=k-x\}\}\\ P\{X+Y=k\}&=\sum_x P(X=x,Y=k-x)\\ &=\sum_x p_{X,Y}(x,k-x).\qquad\text{(Joint pmf } p_{X,Y})\end{aligned}$$

*If at this point $X, Y$ are independent,*

$$P\{X+Y=k\} = \sum_x p_X(x) p_Y(k-x)$$
$$P_{X+Y}(k) := \sum_y p_X(k-y) p_Y(y).$$

*We call this sum the convolution of $X, Y$.*

**Example 6.13.** Let $X \sim \text{Bin}(n_1, p)$. $Y \sim \text{Bin}(n_2, p)$ be independent. Then

$$\begin{aligned}
P\{X+Y=k\} &= \sum_{i=0}^{\infty} P(X=i, Y=k-i) \\
&= \sum_{i=0}^{k} P(X=i) P(Y=k-i) \\
&= \sum_{i=0}^{k} \binom{n_1}{i} p^i q^{n_1-i} \binom{n_2}{k-i} p^{k-i} q^{n_2-(k-i)} \\
&= \sum_{i=0}^{k} \binom{n_1}{i} p^k q^{n_1+n_2-k} \binom{n_2}{k-i} \\
&= \sum_{i=0}^{k} \binom{n_1}{i} p^k q^{n_1+n_2-k} \binom{n_2}{k-i} \\
&= p^k q^{n_1+n_2-k} \sum_{i=0}^{k} \binom{n_1}{i} \binom{n_2}{k-i} \\
&= p^k q^{n_1+n_2-k} \binom{n_1+n_2}{k}. \qquad \text{(Vandermonde)}
\end{aligned}$$

This implies that $X+Y \sim \text{Bin}(n_1+n_2, p)$.

**Note.** This ends the material for Exam 1.

## 6.2 Lecture 10. Wed Oct 2

### 6.2.1 Conditional Distributions

**Definition 6.14.** The conditional distribution of $Y$ given $X=x$ is defined by

$$F_{Y|X}(y \mid x) = P(Y \le y \mid X = x),$$

whenever $P(X=x) > 0$. The conditional probability mass function of $Y$ given $X=x$ is

$$p_{Y|X} = P(Y=y \mid X=x).$$

We can also write this

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

**Note.** To convince ourselves this actually defines a pmf, consider

$$\begin{aligned}\sum_y p_{Y|X}(y \mid x) &= \sum_y \frac{p_{X,Y}(x,y)}{p_X(x)} \\ &= \frac{1}{p_X(x)} \sum_y p_{X,Y}(x,y) \\ &= \frac{1}{p_X(x)} p_X(x) \\ &= 1.\end{aligned}$$

**Example 6.15.** Consider

| $X \backslash Y$ | 1 | 2 |
|---|---|---|
| 0 | 0.1 | 0.2 |
| 1 | 0.5 | 0.2 |

.

Then

$$p_{Y|X}(2 \mid 0) = \frac{p_{X,Y}(0,2)}{p_X(0)} = \frac{0.2}{0.2+0.1} = \frac{2}{3},$$

and

$$p_{X|Y}(1 \mid 1) = \frac{P_{X,Y}(1,1)}{p_Y(1)} = \frac{0.5}{0.1+0.5} = \frac{5}{6}.$$

**Example 6.16** (Quiz 3 continued)**.** Let $X_1 \sim \text{Poisson}(\lambda_1), X_2 \sim \text{Poisson}(\lambda_2)$ be independent. Then $X_1+X_2 \sim \text{Poisson}(\lambda_1+\lambda_2)$. Find the distribution of $X \mid X+Y$, that is, find the pmf of $X \mid X+Y = n$

*Proof.* We want to find $P(X = k \mid X + Y = n) = p_{X|X+Y}(k \mid n)$. Through some algebra,

$$\begin{aligned}P(X = k \mid X + Y = n) &= \frac{P(X = k, X + Y = n)}{P(X + Y = n)} \\ &= \frac{P(X = k, Y = n - k)}{P(X + Y = n)} \\ &= \frac{P(X = k)P(Y = n - k)}{P(X + Y = n)} \\ &= \frac{\dfrac{e^{-\lambda_1}\lambda_1^k}{k!} \dfrac{e^{-\lambda_2}\lambda_2^{n-k}}{(n-k)!}}{\dfrac{e^{-(\lambda_1+\lambda_2)}(\lambda_1+\lambda_2)^n}{n!}} \\ &= \frac{n!}{k!(n-k)!} \cdot \frac{\lambda_1^k \lambda_2^{n-k}}{(\lambda_1+\lambda_2)^k(\lambda_1+\lambda_2)^{n-k}} \\ &= \binom{n}{k}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^k\left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^{n-k} \\ &= \binom{n}{k}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^k\left(1-\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^{n-k}\end{aligned}$$

□

**Exercise 6.1.** Let $X, Y \sim \text{Geom}(p)$ be independent. Find the conditional pmf of $X$ given $X+Y = n$.

*Proof.* Recall that $P(X+Y=n) = (n-1)p^2(1-p)^{n-2}$ (in class exercise). Then

$$\begin{aligned} P(X=k \mid X+Y=n) &= \frac{P(X=k)P(Y=n-k)}{P(X+Y=n)} \\ &= \frac{(1-p)^{k-1}p \cdot (1-p)^{n-k-1}p}{(n-1)p^2(1-p)^{n-2}} \\ &= \frac{1}{n-1}. \end{aligned}$$

□

**Definition 6.17** (Discrete uniform distributions)**.** Let $X$ be a random variable, such that $X \to \{1, 2, \ldots, n\}$, and $P(X=k) = \frac{1}{n}$, then $X$ is a discrete uniformly distributed random variable.

# 7 Week 7

## 7.1 Lecture 11. Mon Oct 7

### 7.1.1 Conditional Expectation

**Note.** Recall $p_{Y|X}(y \mid x) = \frac{p_{X,Y}(x,y)}{p_X(x)} = P(Y=y \mid X=x)$.

**Definition 7.1.** The conditional expectation of $Y$ when $X=x$ is defined by

$$\mathbf{E}[Y \mid X=x] = \sum_y y \cdot p_{Y|X}(y \mid x).$$

**Proposition 7.2.** *We know $\mathbf{E}[Y]$ is a constant. However, $\mathbf{E}[Y \mid X=x]$ is a function of $x$. Let $\psi(X) = \mathbf{E}[Y \mid X]$. We call $\psi(X)$ the conditional expectation of $Y$ given $X$. Note that $\psi(X)$ is a random variable.*

**Theorem 7.3.** *Consider $\psi(X) = \mathbf{E}[Y \mid X]$. Then $\mathbf{E}[\psi(X)] = \mathbf{E}[\mathbf{E}[Y \mid X]] = \mathbf{E}[Y]$.*

*Proof.* We know

$$\begin{aligned} \mathbf{E}[\psi(X)] &= \sum_x \psi(x) p_X(x) \\ &= \sum_x \mathbf{E}[Y \mid X=x]\, p_X(x) \\ &= \sum_x \sum_y y \cdot p_{Y|X}(y \mid x) \cdot p_X(x) \\ &= \sum_x \sum_y y \cdot \frac{P(x,y)}{p_X(x)} p_X(x) \\ &= \sum_y y \sum_x p(x,y) \\ &= \sum_y p_Y(y) = \mathbf{E}[Y]. \end{aligned}$$

□

**Example 7.4.** Consider

$$\mathbf{E}[X] = \sum_y \mathbf{E}[X \mid Y = y]\, P(Y = y) = \mathbf{E}[X \mid B] \cdot P(B) + \mathbf{E}\left[X \mid B^C\right] P(B^C).$$

**Example 7.5.** Consider

| $X \backslash Y$ | 1 | 2 |
|---|---|---|
| 2 | 0.2 | 0.3 |
| 3 | 0.3 | 0.5 |

.

Then

$$\begin{aligned}
\mathbf{E}[X \mid Y = 1] &= \sum_x x \cdot p_{X|Y}(x \mid y) \\
&= 2p_{X|Y}(2 \mid 1) + 3p_{X|Y}(3 \mid 1) \\
&= 2 \cdot \frac{P(2,1)}{p_Y(1)} + 3 \cdot \frac{P(3,1)}{p_Y(1)} \\
&= 2\frac{0.2}{0.5} + 3\frac{0.3}{0.5} = \frac{1.3}{0.5} = 2.6.
\end{aligned}$$

Also,

$$\begin{aligned}
\mathbf{E}[X \mid Y = 2] &= 2p_{X|Y}(2 \mid 2) + 3p_{X|Y}(3 \mid 2) \\
&= 2\frac{P(2,2)}{p_Y(2)} + 3\frac{P(3,2)}{p_Y(2)} \\
&= 2\frac{0.3}{0.5} + 3\frac{0.2}{0.5} = 2.4.
\end{aligned}$$

We can also find

$$\mathbf{E}[X] = \mathbf{E}[X \mid Y = 1]\, P(Y = 1) + \mathbf{E}[X \mid Y = 2]\, P(Y = 2) = (2.6)(0.5) + (2.4)(0.5) = 2.5.$$

**Theorem 7.6.**
*1. The conditional expected value* $\mathbf{E}[k \mid X] = k$.
*2. Conditional expectation is linear:*

$$\mathbf{E}[aX + bY \mid Z] = a\,\mathbf{E}[X \mid Z] + b\,\mathbf{E}[Y \mid Z].$$

*3. If* $X, Y$ *are independent, then* $\mathbf{E}[X \mid Y] = \mathbf{E}[X]$.
*4. For any real valued function* $g$, $\mathbf{E}[g(Y)X \mid Y] = g(Y)\,\mathbf{E}[X \mid Y]$.
*5. The tower property / law of iterated expectation is as follows:*

$$\mathbf{E}[\mathbf{E}[X \mid Y, Z] \mid Y] = \mathbf{E}[X \mid Y].$$

*Proof of 3.* If $X, Y$ are independent, then

$$\begin{aligned}
\mathbf{E}[X \mid Y = y] &= \sum_x x \cdot p_{X|Y}(x \mid y) \\
&= \sum_x x \cdot \frac{P(x,y)}{p_Y(y)} \\
&= \sum_x x \cdot \frac{p_X(x)p_Y(y)}{p_Y(y)} \\
&= \mathbf{E}[X].
\end{aligned}$$

□

*Proof of 4.* Let $g$ be a real valued function. Then

$$\begin{aligned}\mathbf{E}\left[g(Y)X \mid Y=y\right] &= \sum_x g(y)x \cdot p_{x|y}(x \mid y)\\ &= g(y)\sum_x x \cdot p_{x|y}(x \mid y)\\ &= g(y)\,\mathbf{E}\left[X \mid Y=y\right].\end{aligned}$$

□

**Example 7.7.** A fair coin is flipped continuously. What is the expected number of flips required to observe the sequence $HH$?

**Proposition 7.8.** *For $X$ a random variable,*

$$\operatorname*{argmin}_a \mathbf{E}\left[X-a\right]^2 = \mathbf{E}\left[X\right],$$

*and*

$$\min_a \mathbf{E}\left[X-a\right]^2 = \mathbf{Var}\left[X\right].$$

**Theorem 7.9.** *If $h$ is a function of $Y$ and $\mathbf{E}\left[h(Y)^2\right] < \infty$, then $\mathbf{E}\left[(X-h(Y)^2)\right] \geq \mathbf{E}\left[(X-\mathbf{E}\left[X \mid Y\right])^2\right]$. In addition, if $\mathbf{E}\left[(X-h(Y))^2\right] = \mathbf{E}\left[(X-\mathbf{E}\left[X \mid Y\right])^2\right]$, then $\mathbf{E}\left[(h(Y)-\mathbf{E}\left[X \mid Y\right])^2\right] = 0$.*

## 7.2 Lecture 12. Wed Oct 9

*Proof of Theorem 7.9.* We have

$$\begin{aligned}\mathbf{E}\left[(X-h(Y))^2\right] &= \mathbf{E}\left[(X-\mathbf{E}\left[X \mid Y\right]+\mathbf{E}\left[X \mid Y\right]-h(Y))^2\right]\\ &= \mathbf{E}\left[((X-\mathbf{E}\left[X \mid Y\right])+(\mathbf{E}\left[X \mid Y\right]-h(Y))^2\right]\\ &= \mathbf{E}\left[(X-\mathbf{E}\left[X \mid Y\right])^2\right]+\overbrace{\mathbf{E}\left[(\mathbf{E}\left[X \mid Y\right]-h(Y))^2\right]}^{\geq 0}\\ &\quad + 2\underbrace{\mathbf{E}\left[(X-\mathbf{E}\left[X \mid Y\right])(\mathbf{E}\left[X \mid Y\right]-h(Y))\right]}_{I}.\end{aligned}$$

Now,

$$\begin{aligned}I &= \mathbf{E}\left[(X-\mathbf{E}\left[X \mid Y\right])(\mathbf{E}\left[X \mid Y-h(Y)\right])\right]\\ &= \mathbf{E}\left[\mathbf{E}\left[(X-\mathbf{E}\left[X \mid Y\right])(\mathbf{E}\left[X \mid Y\right]-h(Y)) \mid Y\right]\right].\end{aligned}$$

Recall that

1. $\mathbf{E}\left[X\right] = \mathbf{E}\left[\mathbf{E}\left[X \mid Y\right]\right]$,
2. $\mathbf{E}\left[g(Y)X \mid Y\right] = g(Y)\,\mathbf{E}\left[X \mid Y\right]$,
3. $\mathbf{E}\left[X \mid Y\right]$ is a function of $Y$.

Therefore,

$$\begin{aligned} I &= \mathbf{E}\left[\underbrace{(\mathbf{E}\left[X \mid Y\right] - h(Y))}_{g(Y)} \underbrace{\mathbf{E}\left[X - \mathbf{E}\left[X \mid Y\right] \mid Y\right]}_{\mathbf{E}[X|Y] - \mathbf{E}[\mathbf{E}[X|Y]|Y]}\right] \\ &= 0 \text{ (Why?)}. \end{aligned}$$

Now,

$$\begin{aligned} \mathbf{E}\left[(X - h(Y))^2\right] &= \mathbf{E}\left[(X - \mathbf{E}\left[X \mid Y\right])^2\right] + \mathbf{E}\left[(\mathbf{E}\left[(\right] X \mid Y) - h(Y))^2\right] \\ &\geq \mathbf{E}\left[(X - \mathbf{E}\left[X \mid Y\right])^2\right]. \end{aligned}$$

□

**Note.** The second statement of Theorem 7.9 states that a certain set of functions of $Y$, $\mathbf{E}\left[X \mid Y\right]$ are the "best" approximations of $X$. (This set of functions is the set of $L^2(Y)$ functions.)

### 7.2.1 Continuous Random Variables

**Definition 7.10.** We call $X$ a continuous random variable if $F(x) = P(X \leq x)$ is a continuous function.

**Definition 7.11.** We call $X$ an absolutely continuous random variable if there exists an integrable function such that $F(x) = \int_{-\infty}^{x} f(y)\, dy$.

**Note.** We will refer to absolutely continuous random variables as just continuous in this course. However, there are continuous random variables that are not absolutely continuous. For example, Cantor random variables are continuous but not absolutely continuous.

**Proposition 7.12.** *Consider $X$ a continuous random variables. Then*

$$P(X \in A) = \int_A f(x)\, dx,$$

*for any $A$ such that $X^{-1}(A) \in \mathcal{F}$. Particularly useful is the fact that*

$$P(a \leq X \leq b) = \int_a^b f(x)\, dx.$$

*Note that $f(x)$ is not a probability.*

**Proposition 7.13.** *In the case that $X$ is a continuous random variable,*

1. $f(x) \geq 0$
2. $\int_{-\infty}^{\infty} f(x)\, dx = 1.$

**Theorem 7.14.** *If $F$ is continuous, then $P(X = x) = 0$, for all $x$.*

*Proof.* Let $y \in \mathbb{R}$. We defined $P(a \leq X \leq b) = \int_a^b f(x)\,dx$. We want to find $P(y \leq X \leq y)$. This is

$$\begin{aligned} P(y \leq X \leq y) &= \int_y^y f(x)\,dx \\ &= F(y) - F(y) \\ &= 0. \end{aligned}$$

□

### 7.2.2 Continuous Distributions

**Example 7.15.** Let $X \sim U(a, b)$ be a uniformly distributed random variable. That is,

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

**Definition 7.16.** Let $X$ be a random variable with distribution function

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0, \lambda \geq 0 \\ 0, \text{otherwise.} \end{cases}$$

Then $X$ has an exponential distribution.

**Definition 7.17** (Normal distribution preview)**.** If $X$ is a random variable with the distribution function

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad x \in \mathbb{R}, -\infty < \mu < \infty, \sigma > 0,$$

then $X$ is a normally distributed random variable. We write $X \sim N(\mu, \sigma^2)$.

### 7.2.3 Expected Value of Continuous Random Variables

**Definition 7.18.** The expected value of a continuous random variable is

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

when this integral exists.

**Theorem 7.19.** *Let $f$ be the pdf of $X$ and $f(x) = 0$, when $x < 0$. Then*

$$\mathbf{E}[X] = \int_0^{\infty} (1 - F(x))\,dx = \int_0^{\infty} P(X > x)\,dx.$$

*Proof.* We have

$$\begin{aligned}\int_0^\infty P(X>x)\,dx &= \int_{x=0}^\infty \int_{y=x}^\infty f(y)\,dy\,dx \\ &= \int_{y=0}^\infty \int_{x=0}^\infty f(y)\,dx\,dy \\ &= \int_{y=0}^\infty f(y)\int_{x=0}^y dy\,dx \\ &= \int_{y=0}^\infty yf(y)\,dy = \int_{-\infty}^\infty yf(y)\,dy \\ &= \mathbf{E}\left[X\right].\end{aligned}$$

□

**Example 7.20.** Suppose the pdf of $X$ is

$$f(x) = \begin{cases} 2x & 0<x<1 \\ 0 & \text{otherwise.}\end{cases}$$

Find $P(X>\frac{1}{2})$, and $\mathbf{E}\left[X\right]$.

# 8 Week 9

## 8.1 Lecture 13. Mon Oct 21

### 8.1.1 Standard Normal Distribution

One distribution worthy of study is the standard normal distribution, which is the normal distribution with parameters $\mu=0$, and $\sigma^2=1$.

**Example 8.1.** Let $f(x)=\frac{1}{\sqrt{2\pi}}e^{\frac{-x^2}{2}}$. Suppose

$$H_n(x) = \frac{(-1)^n d^n f(x)}{dx^n} = H_n(x)f(x).$$

For instance,

$$H_1(x) = \frac{1}{\sqrt{2\pi}}\cdot\frac{-2x}{2}e^{\frac{-x^2}{2}} = -xf(x).$$

Consider

$$\begin{aligned}H_{n+1}(x)f(x) &= (-1)^{n+1}\frac{d^{n+1}f(x)}{dx^{n+1}} \\ &= -\left[\frac{d}{dx}\left(\frac{(-1)^n d^n f(x)}{dx^n}\right)\right] \\ &= -\frac{d}{dx}[H_n(x)f(x)] \\ &= -[H_n'(x)f(x)+H_n(x)f'(x)] \\ &= -[H_n'(x)f(x)-H_n(x)xf(x)] \\ \Rightarrow H_{n+1}(x) &= xH_n(x)-H_n'(x). \qquad (f(x)\text{ is positive})\end{aligned}$$

**Note.** These polynomials $H_n(x)$ are called Hermite polynomials.

**Example 8.2.** Let $f(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$. Let $\Phi(x) = \int_{-\infty}^{x} f(u)\,du$. Prove that

$$\frac{1-\Phi(x)}{f(x)} \leq \frac{1}{x}, \quad \text{if } x > 0.$$

*Solution.* Note that

$$1 - \Phi(x) = P(X > x) \int_x^\infty f(u)\,du \leq \int_x^\infty \frac{u}{x} f(u)\,du.$$

Now,

$$\begin{aligned} 1 - \Phi(x) &\leq \frac{1}{x} \int_x^\infty u f(u)\,du \\ &= \frac{-1}{x} \int_x^\infty f'(u)\,du \qquad \text{(From previous problem)} \\ &= \frac{-1}{x}(0 - f(x)) \\ &= \frac{f(x)}{x}. \end{aligned}$$

□

**Example 8.3.** Suppose $X \sim U(0,1)$, and $Y = e^X$. Find $f_Y(y), F_Y(y)$.

*Solution.* Take logs of both sides of $P[e^X \leq y]$. □

### 8.1.2 Joint Continuous Random Variables

**Definition 8.4.** The joint distribution of $X, Y$ continuous random variables is given by

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

where $f$ is the joint density function of $X, Y$.

**Definition 8.5.** Consider

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

and

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

These are called marginal density functions of $X, Y$.

**Proposition 8.6.** *We have, for $X, Y$ continuous random variables,*

1. $f(x,y) \geq 0$, $\forall x, y$,
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y)\,dy\,dx = 1.$

**Example 8.7.**
1. The joint density function of $X, Y$ continuous random variables is given by

$$f(x,y) = \begin{cases} k(1-y) & 0 \le x \le y \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

Find $k$, and find $f_X(x), f_Y(y)$.

*Solution.* We know

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dy\,dx \\ &= \int_{x=0}^{1}\int_{y=x}^{1} k(1-y)\,dy\,dx \\ &= k\int_{x=0}^{1}\left[y - \frac{y^2}{2}\right]_x^1 dx \\ &= k\int_{x=0}^{1}\left[1-\frac{1}{2}\right] - \left[x - \frac{x^2}{2}\right] dx \\ &= k\int_{x=0}^{1}\left[\frac{1}{2} - x + \frac{x^2}{2}\right] dx \\ &= k\left[\frac{1}{2}x - \frac{x^2}{2} + \frac{x^3}{6}\right]_0^1 \\ &= k\left(\frac{1}{6}\right) \\ \Rightarrow k &= 6. \end{aligned}$$

Now,

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f(x,y)\,dy \\ &= \int_{y=x}^{1} 6(1-y)\,dy \\ &= 6\left[y - \frac{y^2}{2}\right]_x^1 \\ &= 6\left[\left(1-\frac{1}{2}\right) - \left(x - \frac{x^2}{2}\right)\right] \\ &= 6\left(\frac{1}{2} - x + \frac{x^2}{2}\right). \qquad (0 \le x \le 1) \end{aligned}$$

Also,

$$\begin{aligned} f_Y(y) &= \int_{x=0}^{x=y} 6(1-y)\,dx \\ &= \left[6(1-y)\cdot x\right]_0^y \\ \Rightarrow f_Y(y) &= 6y(1-y). \qquad (0 \le y \le 1) \end{aligned}$$

□

2. For sufficiently nice sets $B \in \mathbb{R}^2$,

$$P((X,Y) \in B) = \int \int_B f(x,y)\, dy\, dx.$$

Find $P(X + Y < 1)$.

## 8.2 Lecture 14. Wed Oct 23

**Definition 8.8.** The random variables $X, Y$ are independent if

$$P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y), \forall x, y.$$

This is equivalent to

$$F(x,y) = F_X(x) \cdot F_Y(y), \forall x, y.$$

**Theorem 8.9.** *The random variables $X, Y$ are independent if and only if*

$$f(x,y) = f_X(x) \cdot f_Y(y), \forall x, y.$$

### 8.2.1 Sums of Continuous Random Variables

**Theorem 8.10.** *Let $Z = X + Y$, for some random variables $X, Y$. Let $f(x,y)$ be the joint density function of $X, Y$. Then*

$$f_Z(z) = \int_{-\infty}^{\infty} f(x, z-x)\, dx.$$

*If $X, Y$ are independent, then*

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\, dx.$$

*Proof.* Consider

$$\begin{aligned} F_Z(z) &= P(Z \leq z) \\ &= P(X + Y \leq z) \\ &= \int \int_A f(x,y)\, dx\, dy \\ &= \int_{x=-\infty}^{\infty} \int_{y=-\infty}^{y=z-x} f(x,y)\, dy\, dx \end{aligned}$$

Let $y = w - x \Rightarrow w = y + x$. Then

$$\int_{x=-\infty}^{\infty} \int_{y=-\infty}^{y=z-x} f(x,y)\, dy\, dx = \int_{x=-\infty}^{\infty} \int_{w=-\infty}^{w=z} f(x, w-x)\, dw\, dx.$$

Recall that $f_Z(z) = \frac{d}{dz} F_Z(z)$. Then taking the derivative of both sides w.r.t. $z$ gives us

$$f_Z(z) = \int_{-\infty}^{\infty} f(x, z-x)\, dx.$$

□

**Example 8.11.** Let $X, Y \sim N(0,1)$ be independent. Set $Z := X + Y$. Find $f_Z(z)$.

*Solution.* Note that $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}$. Then

$$\begin{aligned}
f_Z(z) &= \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\, dx \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{(z-x)^2}{2}}\, dx \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-\frac{1}{2}[x^2 + (z^2 - 2zx + x^2)]}\, dx \\
&= \frac{1}{2\pi} e^{-\frac{z^2}{2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}[2x^2 - 2zx]}\, dx \\
&= \frac{1}{2\pi} e^{-\frac{z^2}{2}} \int_{-\infty}^{\infty} e^{-(x^2 - zx)}\, dx \\
&= \frac{1}{2\pi} e^{-\frac{z^2}{2}} \int_{-\infty}^{\infty} e^{-(x^2 - \frac{z}{2})^2 + \frac{z^2}{4}}\, dx \\
&= \frac{1}{2\pi} e^{-\frac{z^2}{2}} e^{\frac{z^2}{4}} \int_{-\infty}^{\infty} e^{-(x^2 - \frac{z}{2})^2}\, dx
\end{aligned}$$

Let $u = (x - \frac{z}{2})\sqrt{2}$. Then

$$\begin{aligned}
\frac{1}{2\pi} e^{\frac{-z^2}{2}} e^{\frac{z^2}{4}} \int_{-\infty}^{\infty} e^{-(x^2 - \frac{z}{2})^2}\, dx &= \frac{1}{2\pi} e^{\frac{-z^2}{4}} \int_{-\infty}^{\infty} e^{-\frac{u^2}{2}} \cdot \frac{du}{\sqrt{2}} \\
&= \frac{1}{2\pi} e^{\frac{-z^2}{4}} \frac{1}{\sqrt{2}} \sqrt{2\pi} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}\, du \\
&= \frac{1}{2\sqrt{\pi}} e^{-\frac{z^2}{4}} \\
&= \frac{1}{\sqrt{2\pi}\sqrt{2}} e^{-\frac{z^2}{2(2)}} \\
&\iff Z \sim N(0,2).
\end{aligned}$$

□

### 8.2.2 Continuous Conditional Distributions

**Definition 8.12.** The distribution function of $X$ given $Y = y$ is defined by

$$F_{X|Y}(\cdot \mid y),$$

and is defined by

$$\begin{aligned}
F_{X|Y}(x \mid y) &= P(X \le x \mid Y = y) \\
&= \int_{-\infty}^{x} \frac{f(u,y)}{f_Y(y)}\, du,
\end{aligned}$$

for any $y$ such that $f_Y(y) > 0$. We also may write

$$f_{X|Y}(x \mid y) = \frac{f(x,y)}{f_Y(y)},$$

so that

$$F_{X|Y}(x \mid y) = \int_{-\infty}^{x} f_{X|Y}(u \mid y)\, du.$$

We call $f_{X|Y}$ the conditional density function of $X \mid Y = y$.

**Proposition 8.13.** *The function $f_{X|Y}(\cdot \mid y)$ is a valid density function.*

*Proof.* That is,

$$\begin{aligned}\int_{-\infty}^{\infty} f_{X|Y}(x \mid y)\, dx &= \int_{-\infty}^{\infty} \frac{f(x,y)}{f_Y(y)}\, dx \\ &= \frac{1}{f_Y(y)} \int_{-\infty}^{\infty} f(x,y)\, dx \\ &= 1.\end{aligned}$$

□

**Example 8.14.** Let

$$f(x,y) = \begin{cases} xy & 0 \le x \le 1, 0 \le y \le 2 \\ 0 & \text{otherwise.} \end{cases}$$

Find $P(X \le \frac{1}{2} \mid Y = \frac{1}{4})$.

*Solution.* We want to find $f_{X|Y}(x \mid y) = \frac{f(x,y)}{f_Y(y)}$. Consider that

$$\begin{aligned} f_Y(y) &= \int_{x=-\infty}^{x=\infty} f(x,y)\, dx \\ &= \int_0^1 xy\, dx \\ &= \frac{y}{2}. \end{aligned}$$

Therefore,

$$f_{X|Y}(x \mid y) = \frac{xy}{\frac{y}{2}} = 2x.$$

(In this case, this implies $X, Y$ are independent.) Hence

$$\begin{aligned} P(X \le \frac{1}{2} \mid Y = \frac{1}{4}) &= \int_{x=0}^{x=\frac{1}{2}} f_{X|Y}(x \mid \frac{1}{4})\, dx \\ &= \int_0^{\frac{1}{2}} 2x\, dx = \frac{1}{4}. \end{aligned}$$

□

# 9 Week 10

## 9.1 Lecture 15. Mon Oct 28

### 9.1.1 Conditional Expectation of Continuous Random Variables

**Definition 9.1.** The conditional expectation of $X$ given $Y = y$ is defined by

$$\mathbf{E}[X \mid Y = y] = \int_{-\infty}^{\infty} x f_{X|Y}(x \mid y)\, dx,$$

where $f_{X|Y}(x \mid y) = \frac{f(x,y)}{f_Y(y)}$.

**Example 9.2.** Suppose

$$f(x,y) = \begin{cases} 6(1-y) & 0 < x < y < 1 \\ 0 & \text{otherwise.} \end{cases}$$

1. Find $\mathbf{E}[X \mid Y = y]$.

   *Solution.* We know

   $$f_Y(y) = \int_{x=0}^{x=y} 6(1-y)\, dx = 6y(1-y),$$

   and

   $$f_{X|Y}(x \mid y) = \frac{6(1-y)}{6y(1-y)} = \frac{1}{y}.$$

   Thus,

   $$\mathbf{E}[X \mid Y = y] = \int_{x=0}^{x=y} x\frac{1}{y}\, dx = \frac{y}{2},$$

   for $0 < y < 1$. □

2. Find $\mathbf{E}[X]$.

   *Solution.* Since $\mathbf{E}[X] = \mathbf{E}[\mathbf{E}[X \mid Y]]$, we have

   $$\begin{aligned} \mathbf{E}[\mathbf{E}[X \mid Y]] &= \int_{y=-\infty}^{y=\infty} \mathbf{E}[X \mid Y = y]\, f_Y(y)\, dy \\ &= \int_{y=0}^{y=1} \frac{y}{2} 6y(1-y)\, dy \\ &= 3\int_0^1 (y^2 - y^3)\, dy \\ &= \frac{1}{4.} \end{aligned}$$

   □

**Proposition 9.3.** *Suppose $X$ is a continuous random variable. Then*

$$\mathbf{E}[X] = \int_{-\infty}^{\infty} \mathbf{E}[X \mid Y = y]\, f_Y(y)\, dy,$$

*and*

$$\mathbf{E}[Y] = \int_{-\infty}^{\infty} \mathbf{E}[Y \mid X = x]\, f_X(x)\, dx.$$

**Exercise 9.1.** Let

$$f(x,y) = \begin{cases} 30xy^2 & x-1 \le y \le 1-x, 0 \le x \le 1 \\ 0 & \text{otherwise.} \end{cases}$$

Find $\mathbf{E}\left[Y \mid X = \frac{1}{2}\right]$.

*Solution.* The final answer is $\mathbf{E}\left[Y \mid X = \frac{1}{2}\right] = 0$. □

**Example 9.4.** Let $X \sim B(n, Y)$, $Y \sim U(0,1)$. Recall that if $X \sim B(n,p)$, then $\mathbf{E}[X] = np$, $\mathbf{Var}[X] = np(1-p)$, and if $Y \sim U(a,b)$, then $\mathbf{E}[Y] = \frac{a+b}{2}$, $\mathbf{Var}[Y] = \frac{(b-a)^2}{12}$. Find

1. $\mathbf{E}[X]$,
2. $\mathbf{E}[XY]$,
3. $\mathbf{Cov}[Y, \mathbf{E}[X \mid Y]]$.

*Solution.* Note that $\mathbf{E}[X] = \mathbf{E}[\mathbf{E}[X \mid Y]] = \mathbf{E}[nY] = n\,\mathbf{E}[Y] = \frac{1}{2}n$. Also,

$$\begin{aligned} \mathbf{E}[XY] &= \mathbf{E}[\mathbf{E}[XY \mid Y]] \\ &= \mathbf{E}[Y\,\mathbf{E}[X \mid Y]] \\ &= n\,\mathbf{E}\left[Y^2\right] \\ &= n(\mathbf{Var}[Y] + (\mathbf{E}[Y])^2) \\ &= n\left(\frac{1}{12} + \frac{1}{4}\right) = \frac{n}{3}. \end{aligned}$$

Finally, recall that $\mathbf{Cov}[X,Y] = \mathbf{E}[XY] - \mathbf{E}[X]\,\mathbf{E}[Y]$. Now,

$$\begin{aligned} \mathbf{Cov}[Y, \mathbf{E}[X \mid Y]] &= \mathbf{Cov}[Y, nY] \\ &= \mathbf{E}[Y(nY)] - \mathbf{E}[Y]\,\mathbf{E}[nY] \\ &= n\,\mathbf{E}\left[Y^2\right] - n\,\mathbf{E}[Y]^2 \\ &= n\,\mathbf{Var}[Y] \\ &= \frac{n}{12}. \end{aligned}$$

□

### 9.1.2 Well Known Inequalities of Probability

**Theorem 9.5.** *Let $Z$ be a nonnegative random variable. I.e. $f(z) = 0, \forall z < 0$. Then*

$$\mathbf{E}[Z] \ge 0.$$

*Proof.* (Assume $Z$ is continuous.) Then

$$\begin{aligned} \mathbf{E}[Z] &= \int_{-\infty}^{\infty} z f(z)\,dz \\ &= \int_{0}^{\infty} z f(z)\,dz \\ &\ge 0. \end{aligned}$$

□

**Theorem 9.6.** *If $X \leq Y$, then $\mathbf{E}[X] \leq \mathbf{E}[Y]$.*

*Proof.* Let $Z = Y - X \geq 0$. Then

$$\mathbf{E}[Z] = \mathbf{E}[Y - X] \geq 0 \Rightarrow \mathbf{E}[Y] - \mathbf{E}[X] \geq 0 \Rightarrow \mathbf{E}[Y] \geq \mathbf{E}[X].$$

□

**Theorem 9.7** (Rajinda calls this "Basic inequality")**.** *Let $X$ be a random variable, and $h$ a non-negative function. Then $\forall a > 0$,*

$$P(h(X) \geq a) \leq \frac{\mathbf{E}[h(X)]}{a}.$$

*Proof.* Let $a > 0$, and set $A := \{x \mid h(x) \geq a\}$. Let

$$I_A(x) = \begin{cases} 1 & x \in A \Rightarrow h(x) \geq a \\ 0 & x \notin A. \end{cases}$$

Notice that $h(X) - aI_A \geq 0$. Therefore, $h(X) \geq aI_A$, so

$$\mathbf{E}[h(X)] \geq a\,\mathbf{E}[I_A] = aP(A) \Rightarrow \mathbf{E}[h(X)] \geq aP[h(x) \geq a].$$

□

**Theorem 9.8** (Markov's inequality)**.** *The probability*

$$P(|X| \geq a) \leq \frac{\mathbf{E}[|X|]}{a}, \quad a > 0.$$

*Proof.* Follows from Theorem 9.7. □

**Theorem 9.9** (Chebyshev's inequality, probability)**.** *The probability*

$$P(|X| \geq a) \leq \frac{\mathbf{E}\left[X^2\right]}{a^2}, a > 0.$$

*Proof.* Consider that

$$|X| \geq a \iff X^2 \geq a^2.$$

Then

$$P(|X| \geq a) = P(X^2 \geq a^2).$$

Then by Theorem 9.7 (since $h(x) = x^2$),

$$\mathbf{E}[|X| \geq a] \leq \frac{\mathbf{E}\left[X^2\right]}{a^2}$$

□

**Example 9.10.** Prove $P(|X - \mathbf{E}[X]| \geq a) \leq \frac{\mathbf{Var}[X]}{a^2}$.

*Proof.* Let $Y = X - \mathbf{E}[X]$. By Chebyshev's inequality,

$$P(|Y| \geq a) \leq \frac{\mathbf{E}\left[Y^2\right]}{a^2} = \frac{\mathbf{E}\left[(X - \mathbf{E}[X])^2\right]}{a^2} = \frac{\mathbf{Var}[X]}{a^2}.$$

□

## 9.2 Lecture 16. Wed Oct 30

**Example 9.11.** Let $X$ be a continuous random variable such that $\mathbf{Var}[X] = 0$. Prove that $X$ is a constant with probability 1.

*Proof.* We will prove $P(X = \mathbf{E}[X]) = 1$. Define $C_n = \{|X - \mathbf{E}[X]| > \frac{1}{n}\}$. Consider

$$P(C_n) = P\left(|X - \mathbf{E}[X]| > \frac{1}{n}\right) \leq \frac{\mathbf{Var}[X]}{\left(\frac{1}{n}\right)^2} = 0.$$

Notice that $\{C_n\}$ is increasing. Let us consider

$$\begin{aligned} P(X \neq \mathbf{E}[X]) &= P\left(\bigcup_{n=1}^{\infty} C_n\right) \\ &= \lim_{n\to\infty} P(C_n) = 0. \end{aligned}$$

Hence, $P(X = \mathbf{E}[X]) = 1$. □

**Definition 9.12.** Let $g : \mathbb{R} \to \mathbb{R}$. We call $g$ convex if $\forall x, a \in \mathbb{R}$, $\exists \lambda_a \in \mathbb{R}$ such that

$$g(x) \geq g(a) + \lambda_a(x - a).$$

**Example 9.13.** Consider $g(x) = e^x$.

**Theorem 9.14** (Jensen's Inequality)**.** *If $g$ is a convex function, then*

$$\mathbf{E}[g(X)] \geq g(\mathbf{E}[X]).$$

*Proof.* Choose $a = \mathbf{E}[X]$. Then there exists $\lambda_a \in \mathbb{R}$ such that

$$\begin{aligned} g(X) &\geq g(\mathbf{E}[X]) + \lambda_a(x - \mathbf{E}[X]) \\ \mathbf{E}[g(X)] &\geq \mathbf{E}[g(\mathbf{E}[X]) + \lambda_a(x - \mathbf{E}[X])] \\ &= \mathbf{E}[g(\mathbf{E}[X])] + \lambda_a \mathbf{E}[X - \mathbf{E}[X]] \\ &= g(\mathbf{E}[X]) + \lambda_a(\mathbf{E}[X] - \mathbf{E}[X]) \\ &= g(\mathbf{E}[X]). \end{aligned}$$

□

**Theorem 9.15** (Hölder's Inequality)**.** *If $P, Q \geq 1$, and $\frac{1}{P} + \frac{1}{Q} = 1$, then*

$$\mathbf{E}[|XY|] \leq \mathbf{E}[|X|^p]^{\frac{1}{p}} \mathbf{E}[|Y|^q]^{\frac{1}{q}}.$$

**Theorem 9.16** (Minkowski's Inequality)**.** *If $P \geq 1$, then*

$$(\mathbf{E}[X + Y]^p)^{\frac{1}{p}} \leq \mathbf{E}[|X|^p]^{\frac{1}{p}} + \mathbf{E}[|Y|^p]^{\frac{1}{p}}$$

### 9.2.1 Convergence of Random Variables

If $X_1, X_2, \ldots,$ is a sequence of random variables on $(\Omega, \mathcal{F}, P)$, and $X$ is another random variable in the same space, how do we interpret the convergence of of $X_n \to X$ as $n \to \infty$. Recall the epsilon-delta definition of convergence from analysis: If $\{a_n\}$ is a sequence of real numbers, we say $a_n \to a$ as $n \to \infty$ if $\forall \varepsilon > 0$, $\exists N \in \mathbb{Z}^+$ such that $|a_n - a| < \varepsilon$, $\forall n \geq N$.

**Definition 9.17** (Modes of convergence of random variables)**.** The following are conditions which we define for a sequence of random variables to converge. Let $\{X_n\}$ be a sequence of random variables, and let $X$ be a random variable, all in $(\Omega, \mathcal{F}, P)$.

1. Almost sure convergence: we say $X_n \xrightarrow{\text{as}} X$ if
$$P(\omega \in \Omega \mid \lim_{n\to\infty} X_n(\omega) = X(\omega)) = 1.$$

2. The $r$th mean convergence: we say $X_n \xrightarrow{r} X$ if $\forall r \geq 1 \in \mathbb{R}$, and $\mathbf{E}\left[|X_n|^r\right] < \infty$, for all $n \in \mathbb{Z}^+$, and
$$\lim_{n\to\infty} \mathbf{E}\left[|X_n - X|\right]^r = 0.$$

3. Convergence in probability: we say $X_n \xrightarrow{P} X$ if $\forall \varepsilon > 0$,
$$\lim_{n\to\infty} P(|X_n - X| > \varepsilon) = 0.$$

4. Convergence in distribution / law / weak convergence: we say $X_n \xrightarrow{D} X$ if
$$\lim_{n\to\infty} P(X_n \leq x) = P(X \leq x).$$
Equivalently,
$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x), \quad \forall x \text{ such that } F_X \text{ is continuous.}$$

**Note.** These conditions for convergence are not equivalent, but we have some nice properties.

**Proposition 9.18.** *Let $\{X_n\}, X$ be as above. Then*
$$X_n \xrightarrow{as} X \implies X_n \xrightarrow{P} X \implies X_n \xrightarrow{D} X,$$
*and $X_n \xrightarrow{r} X \implies X_n \xrightarrow{P} X$. Any other implications do not hold in general.*

**Example 9.19.** Let $X_n \xrightarrow{1} X$. Prove that $X_n \xrightarrow{P} X$.

*Proof.* Use Markov's inequality. □

# 10 Week 11

## 10.1 Lecture 17

**Theorem 10.1.** *Let $X_n \xrightarrow{P} L$, and let $g$ be continuous at $L$. Then $g(X_n) \xrightarrow{P} L$.*

*Proof.* Let $\varepsilon > 0$ be given. Then there exists $\delta > 0$, such that

$$\begin{aligned}
&|x - L| < \delta \Longrightarrow |g(x) - g(L)| < \varepsilon \\
\Longrightarrow\ &|g(x) - g(L)| \geq L \Longrightarrow |x - L| \geq \delta \\
\Longrightarrow\ &\{|g(X_n) - g(L)| \geq \varepsilon\} \subset \{|X_n - L| \geq \delta\} \\
\Longrightarrow\ &P\{|g(X_n - g(L))| \geq \varepsilon\} \leq P\{|X_n - L| \geq \delta\} \to 0, \text{ as } n \to \infty \\
\Longrightarrow\ &P\{|g(X_n) - g(L)| \geq \varepsilon\} \to 0.
\end{aligned}$$

□

**Lemma 10.2.** *For all $a \in \mathbb{R}$, and for all $\varepsilon > 0$,*

$$P[Y \leq a] \leq P[X \leq a + \varepsilon] + P[|Y - X| > \varepsilon].$$

*Proof.* Consider

$$\begin{aligned}
P[Y \leq a] &= P(Y \leq a, X \leq a + \varepsilon) + P(Y \leq a, X > a + \varepsilon) \\
&\leq P[X \leq a + \varepsilon] + P(Y - X \leq a - X, a - X < -\varepsilon) \\
&\leq P[X \leq a + \varepsilon] + P(Y - X < -\varepsilon) \\
&\leq P[X \leq a + \varepsilon] + P(Y - X < -\varepsilon) + P(Y - X > \varepsilon) \\
&= P[X \leq a + \varepsilon] + P(|Y - X| > \varepsilon).
\end{aligned}$$

□

**Theorem 10.3.** *If $X_n \xrightarrow{P} X$, then $X_n \xrightarrow{D} X$.*

*Proof.* Let $a \in \mathbb{R}$, $\varepsilon > 0$. Then

$$P(X_n \leq a) \leq P(X \leq a + \varepsilon) + P(|X_n - X| > \varepsilon).$$

Consider also

$$P(X \leq a - \varepsilon) \leq P(X_n \leq a) + P(|X_n - X| > \varepsilon).$$

Hence

$$\begin{aligned}
P(X \leq a - \varepsilon) - P(|X_n - X| > \varepsilon) \leq P(X_n \leq a) &\leq P(X \leq a + \varepsilon) + P(|X_n - X| > \varepsilon) \\
\Longrightarrow P(X \leq a - \varepsilon) \leq \lim_{n \to \infty} P(X_n \leq a) &\leq P(X \leq a + \varepsilon) \\
F_X(a - \varepsilon) \leq \lim_{n \to \infty} F_n(a) &\leq F_X(a + \varepsilon).
\end{aligned}$$

If $F_X$ is continuous at $a$ as $\varepsilon \to 0$, then

$$\begin{aligned}
&F_X(a) \leq \lim_{n \to \infty} F_n(a) \leq F_X(a) \\
\Longrightarrow\ &\lim_{n \to \infty} F_n(a) = F_X(a) \\
\Longrightarrow\ &X_n \xrightarrow{D} X.
\end{aligned}$$

□

### 10.1.1 Law of Large Numbers

**Proposition 10.4.** *Let $\{X_n\}$ be i.i.d (independently and identically distributed), with $\mathbf{E}[X_i] = \mu$, $\mathbf{Var}[X_i] = \sigma^2$, and*

$$S_n = \sum_{k=1}^{n} X_k.$$

*Then*

$$\mathbf{E}[S_n] = \sum_{k=1}^{n} \mathbf{E}[X_i] = n\mu.$$

*Also,*

$$\mathbf{Var}[S_n] = \sum_{k=1}^{n} \mathbf{Var}[X_i] = n\sigma^2,$$

*and*

$$\mathbf{Var}\left[\sum_{k=1}^{n} x_i\right] = \sum_{k=1}^{n} \mathbf{Var}[X_i] + 2 \sum_{1 \le i < j \le n} \mathbf{Cov}[X_i, X_j].$$

**Theorem 10.5** (Weak law of large numbers)**.** *Let $\{X_n\}$ be i.i.d, with $\mathbf{E}[X_i] = \mu < \infty$, and $\mathbf{Var}[X_i] < \infty$. Then*

$$\frac{S_n}{n} = \frac{X_1 + \ldots + X_n}{n} \xrightarrow{P} \mu.$$

*Proof.* Let $\varepsilon > 0$. Then we want to show that

$$\lim_{n \to \infty} P\left[\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right] = 0.$$

By Chebyshev's inequality,

$$\begin{aligned} P\left[\left|\frac{S_n}{n} - \mu\right| > \varepsilon\right] &\le \frac{\mathbf{E}\left[(\frac{S_n}{n} - \mu)^2\right]}{\varepsilon^2} \\ &= \frac{\mathbf{E}\left[(S_n - n\mu)^2\right]}{n^2\varepsilon^2} \\ &= \frac{\mathbf{E}\left[(S_n - \mathbf{E}[S_n])^2\right]}{n^2\varepsilon^2} \\ &= \frac{\mathbf{Var}[S_n]}{n^2\varepsilon^2} \\ &= \frac{n\sigma^2}{n^2\varepsilon^2} \\ &= \frac{\sigma^2}{n\varepsilon^2} \to 0 \text{ as } n \to \infty. \end{aligned}$$

□

**Example 10.6.** Let $\{X_n\}$ be i.i.d, with $X_i \sim \mathrm{Bin}(m, p)$. (Thus, $\mathbf{E}[X_i] = mp$). Hence

$$\frac{X_1 + X_2 + \ldots + X_n}{n} \xrightarrow{P} mp \text{ as } n \to \infty.$$

**Theorem 10.7** (Strong law of large numbers)**.** *If $\{X_n\}$ is i.i.d, and $\mathbf{E}[X_i] = \mu < \infty$, then*

$$\frac{X_1 + \ldots + X_n}{n} \xrightarrow{as} \mu.$$

**Note.** In general, if $\{X_n\}$ is a sequence of random variables, and $\mathbf{E}[X_i] = \mu_i < \infty$, we say $\{X_n\}$ obeys the weak law of large numbers if for all $\varepsilon > 0$,

$$P\left[\left|\frac{X_1 + X_2 + \ldots + X_n}{n} - \frac{\mu_1 + \mu_2 + \ldots + \mu_n}{n}\right| > \varepsilon\right] = 0.$$

We say $\{X_n\}$ obeys the strong law of large numbers if almost sure convergence holds.

**Theorem 10.8** (Kolmogorov's first theorem)**.** *If $\{X_n\}$ are i.i.d, then $\{X_n\}$ obeys the weak law of large numbers iff $\mathbf{E}[|X_i|] < \infty$.*

**Theorem 10.9** (Chebyshev's theorem)**.** *If $\{X_n\}$ is a sequence of random variables such that $X_i$ is independent from $X_j$ for every $i \neq j$, and there exists an $M$ such that $\mathbf{Var}[X_i] \leq M$, for all $n \in \mathbb{Z}^+$, then $\{X_n\}$ obeys the weak law of large numbers.*

**Theorem 10.10** (Markov's theorem)**.** *If*

$$\frac{1}{n^2}\mathbf{Var}[X_1 + \ldots + X_n] \to 0,$$

*then $\{X_n\}$ obeys the weak law of large numbers.*

**Theorem 10.11** (Kolmogorov's second theorem)**.** *If $\{X_n\}$ are independent, and $\mathbf{Var}[X_i] = \sigma_i^2$, then*

$$\sum_{i=1}^{\infty} \frac{\sigma_i^2}{n^2} < \infty \Longrightarrow \{X_n\} \text{ obeys the strong law of large numbers.}$$

# 11 Week 12

## 11.1 Lecture 18. Mon Nov 11

### 11.1.1 Moment Generating Functions

**Definition 11.1.** A moment generating function $M$ of $X$ is a function defined as

$$M(t) = \mathbf{E}\left[e^{tX}\right]$$

for all $t$ such that $M(t) < \infty$.

**Proposition 11.2** (An assumption)**.** *We will assume that $M$ is smooth over some ball containing $0$. (That is, $M^{(n)}$ exists for all $n \in \mathbb{Z}^+$.)*

**Theorem 11.3.** *Let $M$ be a moment generating function of $X$. Then $M^{(n)}(0) = \mathbf{E}[X^n]$.*

*Proof.* The Taylor expansion of $M$ centered at $0$ gives

$$M(t) = \sum_{n=0}^{\infty} \frac{M^{(n)}(0) \cdot t^n}{n!}.$$

Also,

$$\begin{aligned}\mathbf{E}\left[e^{tX}\right] &= \mathbf{E}\left[\sum_{n=0}^{\infty}\frac{(tX)^n}{n!}\right] \\ &= \sum_{n=0}^{\infty}\frac{\mathbf{E}\left[X^n\right]t^n}{n!}X \qquad\qquad = M(t).\end{aligned}$$

This implies that $M^{(n)}(0) = \mathbf{E}\left[X^n\right]$. □

**Theorem 11.4.** *If $X_1, \ldots, X_n$ are independent and $M_{X_i}$ denotes the moment generating function of $X_i$, then the moment generating function of the sum*

$$\sum_{i=1}^{n} X_i \quad \textit{is} \quad \prod_{i=1}^{n} M_{X_i}(t).$$

*Or we can write*

$$M_{\sum_{i=1}^n X_i}(t) = \prod_{i=1}^{n} M_{X_i}(t).$$

*Proof.* Write

$$\begin{aligned}M_{\sum_{i=1}^n X_i}(t) &= \mathbf{E}\left[e^{t(\sum_{i=1}^n X_i)}\right] \\ &= \mathbf{E}\left[e^{tX_1}e^{tX_2}\cdots e^{tX_n}\right] \\ &= \mathbf{E}\left[e^{tX_1}\right]\cdot\mathbf{E}\left[e^{tX_2}\right]\cdots\mathbf{E}\left[e^{tX_n}\right] \qquad \text{(since } X_1,\ldots,X_n \text{ are independent)} \\ &= M_{X_1}(t)M_{X_2}(t)\cdots M_{X_n}(t).\end{aligned}$$

□

**Theorem 11.5.** *Let $Y = aX$. Then*

$$M_Y(t) = M_X(at).$$

*Proof.* Consider that

$$M_Y(t) = \mathbf{E}\left[e^{tY}\right] = \mathbf{E}\left[e^{(t\cdot a)x}\right] = M_X(at).$$

□

**Theorem 11.6.** *Let $M_X$, $M_Y$ denote the moment generating functions of $X$, $Y$. If there exists $a > 0$ such that*

$$M_X(t) = M_Y(t), \quad \forall t \in (-a, a),$$

*then $X, Y$ have the same distribution.*

**Theorem 11.7** (Levy's Continuity Theorem)**.** *Let $\{X_n\}$ be a sequence of random variables, and $\{M_n\}$ denote the corresponding moment generating functions. If $M_n(t) \to M_X(t)$, for all $t \in (-a, a)$, then $X_n \xrightarrow{D} X$.*

**Example 11.8.** Let $X \sim N(\mu, \sigma^2)$. Then

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

Consider $\mu = 0$, $\sigma = 1$ ($X \sim N(0,1)$). Then

$$\begin{aligned} M(t) = \mathbf{E}\left[e^{tx}\right] &= \int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}\, dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{-(x^2 - 2tx)}{2}}\, dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{\frac{-[(x-t)^2 - t^2]}{2}}\, dx \\ &= \frac{e^{\frac{t^2}{2}}}{\sqrt{2\pi}} \underbrace{\int_{-\infty}^{\infty} e^{\frac{-(x-t)^2}{2}}\, dx}_{\sqrt{2\pi}} \\ &= e^{\frac{t^2}{2}}. \end{aligned}$$

### 11.1.2 Central Limit Theorem

If $X_1, X_2, \ldots$ are i.i.d, and

$$\overline{X_n} = \frac{\sum_{i=1}^n X_i}{n}, \qquad \text{(called sample mean)}$$

and $\mathbf{E}[X_i] = \mu, \mathbf{Var}[X_i] = \sigma^2$, then $\mathbf{E}\left[\overline{X_n}\right] = \mu$, and $\mathbf{Var}\left[\overline{X_n}\right] = \frac{\sigma^2}{n}$. We want to make some statements about the behavior of $\overline{X_n}$ as $n \to \infty$.

**Theorem 11.9** (Central Limit Theorem)**.** *If $\{X_n\}$ is an i.i.d sequence of random variables, with $\mathbf{E}[X_i] = \mu < \infty$, and $\mathbf{Var}[X_i] = \sigma^2$, then*

$$\frac{\overline{X_n} - \mu}{\frac{\sigma}{\sqrt{n}}} \xrightarrow{D} Z \sim N(0,1).$$

*Or,*

$$\frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma} \xrightarrow{D} N(0,1).$$

*Proof.* Assume $\mu = 0$, $\sigma = 1$, with $M$, the moment generating function of $M$, smooth over $(-a, a)$, for some $a > 0$. We have

$$\frac{\frac{\sum_{i=1}^n X_i}{n} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{\sum_{i=1}^n X_i}{\sqrt{n}}.$$

The moment generating function of $\frac{X_i}{\sqrt{n}}$ is $M(\frac{t}{\sqrt{n}})$. The moment generating function of

$$\frac{\sum_{i=1}^n X_i}{\sqrt{n}} = \left[M\left(\frac{t}{\sqrt{n}}\right)\right]^n = M_{\frac{\sum_{i=1}^n X_i}{\sqrt{n}}}(t).$$

Our goal is to prove

$$M_{\frac{\sum_{i=1}^{n} X_i}{\sqrt{n}}}(t) \to e^{\frac{t^2}{2}}$$
$$\implies \frac{\sum_{i=1}^{n} X_i}{\sqrt{n}} \xrightarrow{D} N(0,1).$$

Set

$$L(t) = \ln M(t).$$

Then

$$L(0) = \ln M(0) = \ln \mathbf{E}\left[e^{0 \cdot x}\right] = \ln 1 = 0.$$

Also,

$$L'(0) = \frac{M'(0)}{M(0)} = \frac{\mathbf{E}\left[X_i\right]}{1} = \frac{\mu}{1} = 0.$$

We also have

$$L''(0) = \frac{M(0)M''(0) - M(0)M''(0)}{M(0)^2} = M''(0) = \mathbf{E}\left[X_i^2\right] = \mathbf{Var}\left[X_i\right] + \mathbf{E}\left[X_i\right]^2$$
$$\implies L''(0) = 1.$$

Now,

$$\begin{aligned}
\lim_{n\to\infty} n \ln M\left(\frac{t}{\sqrt{n}}\right) &= \lim_{n\to\infty} nL\left(\frac{t}{\sqrt{n}}\right) \\
&= \lim_{n\to\infty} \frac{L(\frac{t}{\sqrt{n}})}{\frac{1}{n}} \\
&= \lim_{n\to\infty} \frac{L'(\frac{t}{\sqrt{n}})(-\frac{1}{2}tn^{-\frac{3}{2}})}{-n^{-2}} \\
&= \lim_{n\to\infty} \frac{L'(\frac{t}{\sqrt{n}})t}{2n^{\frac{-1}{2}}} \\
&= \lim_{n\to\infty} \frac{L''(\frac{t}{\sqrt{n}}) \cdot t \cdot t \cdot (-\frac{1}{2}n^{-\frac{3}{2}})}{-n^{\frac{-3}{2}}} \\
&= \frac{1}{2}t^2 \lim_{n\to\infty} L''\left(\frac{t}{\sqrt{n}}\right) \\
&= \frac{1}{2}t^2 L''(0) \\
&= \frac{1}{2}t^2.
\end{aligned}$$

This is sufficient to prove that

$$\lim_{n\to\infty} \left[M\left(\frac{t}{\sqrt{n}}\right)\right]^n = e^{\frac{t^2}{2}}.$$

By Levi's continuity theorem,

$$M_{\frac{\sum_{i=1}^{n} X_i}{\sqrt{n}}}(t) \to e^{\frac{-t^2}{2}} \implies \frac{\sum_{i=1}^{n} x_i}{\sqrt{n}} \xrightarrow{D} N(0,1).$$

□

**Example 11.10.**
1. Suppose $\{X_n\}$ is a sequence of independent random variables such that $X_n \sim B(m,p)$, for $n \in \mathbb{Z}^+$. Then $\mathbf{E}[X_i] = mp$, and $\mathbf{Var}[X_i] = mp(1-p)$. Therefore,

$$\frac{\overline{X_i} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{\overline{X_n} - mp}{\sqrt{\frac{mp(1-p)}{n}}} \xrightarrow{D} N(0,1)$$

This means that for sufficiently large $n$,

$$\frac{\overline{X_n} - mp}{\sqrt{\frac{mp(1-p)}{n}}} \quad \dot{\sim} \quad N(0,1).$$

2. Suppose $\{X_n\}$ is a sequence of independent random variables such that $X_n \sim U(0,1), \forall n \in \mathbb{Z}^+$. Estimate

$$P\bigg(\sum_{i=1}^{100} X_i \leq 55\bigg).$$

Now, since $X_n$ is uniform for all $n \in \mathbb{Z}^+$, $\mathbf{E}[X_n] = \frac{1}{2}, \mathbf{Var}[X_i] = \frac{1}{12}, \mu = \frac{1}{2}, \sigma^2 = \frac{1}{12}$, and $n = 100$. Therefore,

$$\begin{aligned} P\bigg[\sum_{i=1}^{n} X_i \leq 55\bigg] &= P\bigg[\frac{\sum_{i=1}^{100} X_i - 100(\frac{1}{2})}{10 \cdot \sqrt{\frac{1}{12}}} \leq \frac{55 - 100(\frac{1}{2})}{10 \cdot \sqrt{\frac{1}{12}}}\bigg] \\ &\approx P\bigg[Z \leq \frac{55-50}{10\sqrt{\frac{1}{12}}}\bigg] \\ &= \Phi\bigg(\frac{55-50}{10 \cdot \sqrt{\frac{1}{12}}}\bigg). \end{aligned}$$

### 11.1.3 Stochastic Processes

**Definition 11.11** (Stochastic Processes)**.** A stochastic process is a collection of random variables $\{X_t\}_{t \in T}$, where $t$ typically represents time.

**Definition 11.12** (Markov Chains)**.** Let $\{X_n\}$ be a sequence of random variables taking values from a finite set $S = \{1, 2, \ldots, M\}$, where $S$ is called the state space. If

$$P[X_{n+1} = j \mid X_n = i_n, \ldots, X_0 = i_0] = P[X_{n+1} = j \mid X_n = i], \quad \forall i, j \in S,$$

then we call $\{X_n\}$ a Markov chain. This property is called the Markov property. One can understand the Markov property as stating that given $X_n$, the next state $X_{n+1}$ depends only on the current state $X_n$.

**Note.** Markov chains are applied heavily in machine learning, specifically in reinforcement learning. See Markov decision processes.

**Definition 11.13** (Transition probabilities, time homogeneity)**.** Denote $P_{ij} = P(X_1 = j \mid X_0 = i)$. We say $\{X_n\}$ is time homogeneous if

$$P_{ij} = P(X_{n+1} = j \mid X_n = i).$$

## 11.2 Lecture 19. Wed Nov 13

### 11.2.1 Markov Chains

**Proposition 11.14.** *The Markov property is equivalent to the following statement:*

$$P(X_{n+m} = j \mid X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = P(X_{n+m} = j \mid X_n = i), \quad \forall i, j \in S.$$

*Proof.* You can prove this using the law of total probability, and induction. Show that there is no dependence between $X_{n+m}$ and $X_{n-1}, X_{n-2}, \ldots, X_0$ given $X_n$. □

**Exercise 11.1.** If $\{X_n\}$ is a Markov chain, is $\{X_{2n}\}$ a Markov chain?

*Proof.* Set $Y_n = X_{2n}$. Consider

$$\begin{aligned} P(Y_{n+1} = j \mid Y_n = i, Y_{n-1}, \ldots, Y_0) &= P(X_{2n+2} = j \mid X_{2n} = i, X_{2n-2}, \ldots, X_0) \\ &= P(X_{2n+2} = j \mid X_{2n} = i) \\ &= P(Y_{n+1} = j \mid Y_n = i). \end{aligned}$$

Therefore, $\{Y_n\}$ is a Markov chain. □

**Note.** Rajinda writes a standalone $Y_{n-1}$, for example, to mean $Y_{n-1} = i_{n-1}$. The value is inconsequential to the proof, so the abuse of notation is meant to speed the lecture along.

**Exercise 11.2.** Suppose $\{X_n\}$ is a Markov chain. Define $Y_n = (X_n, X_{n+1})$, $n \geq 0$. Prove that $Y_n$ is a Markov chain.

*Proof.* Let $S$ be the state space of $\{X_n\}$. Then $S \times S$ is the state space of $\{Y_n\}$. Then consider

$$\begin{aligned} P[Y_{n+1} = (j,k) \mid Y_n = (i,\ell), Y_{n-1}, \ldots, Y_0] &= P[X_{n+1} = j, X_{n+2} = k \mid X_n = i, X_{n+1} = \ell, X_{n-1}, \ldots, X_0] \\ &= P[X_{n+1} = j, X_{n+2} = k \mid X_n = i, X_{n+1} = \ell] \\ &= P[Y_{n+1} = (j,k) \mid Y_n = (i,\ell)]. \end{aligned}$$

□

### 11.2.2 Transition Probability

Recall Definition 11.13. Namely, for the short duration of the remainder of this course, we will consider time homogeneous Markov chains. Notice that each index $i, j$ is a member of $S$, the state space of $\{X_n\}$, suppose $S = \{1, 2, \ldots, M\}$. Define an $M \times M$ matrix $Q = [P_{ij}]$.

**Definition 11.15.** A matrix $Q$ is called a stochastic matrix if

1. $P_{ij} \geq 0, \ \forall i, j \in S,$
2. $\sum_{j \in S} P_{ij} = 1$, that is, the row sums of $Q$ are all 1.

Additionally, if $\sum_{i \in S} P_{ij} = 1$, then we say $Q$ is doubly stochastic.

**Definition 11.16.** The $n$th transition probability from state $i$ to $j$ is

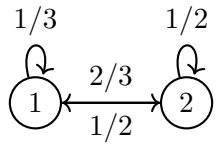
$$P_{ij}^{(n)} = P(X_n = j \mid X_0 = i).$$

**Example 11.17.** Consider the following transition matrix

$$Q = \begin{pmatrix} 1/3 & 2/3 \\ 1/2 & 1/2 \end{pmatrix}.$$

Let $S = \{1, 2\}$. Then

$$Q = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix}.$$




How can we find $P(X_2 = 1 \mid X_0 = 1) = P_{11}^{(2)}$? Write $Q^{(2)} = QQ$ to be the 2-step transition matrix:

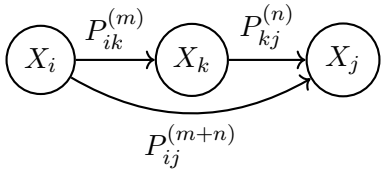
$$Q^{(2)} = QQ = \begin{pmatrix} 1/3 & 2/3 \\ 1/2 & 1/2 \end{pmatrix} \begin{pmatrix} 1/3 & 2/3 \\ 1/2 & 1/2 \end{pmatrix} = \begin{pmatrix} 4/9 & 5/9 \\ 5/12 & 7/12 \end{pmatrix} = \begin{pmatrix} P_{11}^{(2)} & P_{12}^{(2)} \\ P_{21}^{(2)} & P_{22}^{(2)} \end{pmatrix}.$$

**Proposition 11.18.** *In general,*

$$P_{ij}^{(2)} = \sum_{k \in S} P_{ik} P_{kj}.$$

**Theorem 11.19** (Chapman-Kolmogorov equation)**.** *We have*

$$P_{ij}^{(m+n)} = \sum_{k \in S} P_{ik}^{(n)} P_{kj}^{(m)}.$$




**Lemma 11.20.** *Let $A, B, C$ be arbitrary events in a sample space $\Omega$. Then*

$$P(A \cap B \mid C) = P(A \mid B \cap C) P(B \mid C).$$

*Proof.* We have

$$P(A \mid B \cap C) P(B \mid C) = \frac{P(A \cap B \cap C)}{P(B \cap C)} \frac{P(B \cap C)}{P(C)} = \frac{P(A \cap B \cap C)}{P(C)}.$$

□

*Proof.* We have that

$$\begin{aligned}
P_{ij}^{(m+n)} &= P(X_{m+n} = j \mid X_0 = i) \\
&= P\left( \bigcup_{k \in S} \{X_{m+n} = j, X_n = k\} \mid X_0 = i \right) \\
&= \sum_{k \in S} P(X_{m+n} = j, X_n = k \mid X_0 = i) \\
&= \sum_{k \in S} P(X_{m+n} = j \mid X_n = k, X_0 = i) P(X_n = k \mid X_0 = i) && \text{(Lemma 11.20)} \\
&= \sum_{k \in S} P(X_{m+n} = j \mid X_n = k) P(X_n = k \mid X_0 = i) && \text{(Markov property)} \\
&= \sum_{k \in S} P(X_m = j \mid X_0 = k) P(X_n = k \mid X_0 = i) \\
&= \sum_{k \in S} P_{ij}^{(n)} P_{kj}^{(m)}.
\end{aligned}$$

□

**Corollary 11.20.1.** *If $Q$ is a transition matrix,*

$$Q^{(n+m)} = Q^{(m)} Q^{(n)}.$$

**Definition 11.21.** Write $\alpha_i = P(X_0 = i)$. Note that

$$\sum_{i \in S} \alpha_i = 1.$$

Also, $\alpha_j^{(n)} = P(X_n = j)$.

**Exercise 11.3.** Let $S = \{1, 2, \ldots, M\}$. Find a relationship between $\alpha_i$ and $\alpha_j^{(n)}$.

*Proof.* We have

$$\alpha_j^{(n)} = P(X_n = j) = \sum_{i \in S} P(X_n = j \mid X_0 = i) P(X_0 = i)$$

$$\alpha_j^{(n)} = \sum_{i \in S} P_{ij}^{(n)} \cdot \alpha_i.$$

Consider now a row vector

$$\alpha^{(n)} = (\alpha_1^{(n)}, \alpha_2^{(n)}, \ldots, \alpha_M^{(n)}),$$

and

$$\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_M).$$

Therefore, we can write

$$\alpha^{(n)} = \alpha Q^{(n)}.$$

□

**Example 11.22.** Let $S = \{1, 2\}$. Then

$$Q = \begin{pmatrix} 1 & 0 \\ 1/2 & 1/2 \end{pmatrix}.$$

Suppose $P(X_0 = 1) = \frac{1}{4}$, $P(X_0 = 2) = \frac{3}{4}$. Find $P(X_3 = 1)$.

*Proof.* It turns out

$$Q^{(3)} = Q^3 = \begin{pmatrix} 1 & 0 \\ 7/8 & 1/8 \end{pmatrix}.$$

Then $\alpha = (\alpha_1, \alpha_2) = \left(\frac{1}{4}, \frac{3}{4}\right)$. Hence

$$\alpha^{(3)} = \alpha Q^{(3)} = \begin{pmatrix} \frac{1}{4} & \frac{3}{4} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 7/8 & 1/8 \end{pmatrix} = \begin{pmatrix} \frac{29}{32} & \frac{3}{32} \end{pmatrix}.$$

Hence $\alpha_1^{(3)} = P(X_3 = 1) = \frac{29}{32}$, and $\alpha_2^{(3)} = P(X_3 = 2) = \frac{3}{32}$. □

**Definition 11.23.** We say that $Q$ is regular if there exists $n_0 < \infty$ such that $P_{ij}^{(n_0)} > 0,\ \forall i, j \in S$. This means that after $n_0$ steps, we can start at any state and travel to another with positive probability.

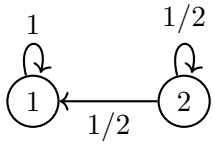
**Example 11.24.** Consider

$$Q = \begin{pmatrix} 0 & 1 \\ 1/2 & 1/2 \end{pmatrix}; \quad Q^2 = \begin{pmatrix} 1/2 & 1/2 \\ 1/4 & 3/4 \end{pmatrix}.$$

Thus, $Q$ is regular. Consider

$$Q' = \begin{pmatrix} 1 & 0 \\ 1/2 & 1/2 \end{pmatrix}; \quad (Q')^n = \begin{pmatrix} * & 0 \\ * & * \end{pmatrix}.$$

That is, $P_{12}^{(n)} = 0, \forall n \in \mathbb{Z}^+$. This means that $Q'$ is not regular.




We call state $1$ an absorbing state.

**Definition 11.25.** We say $Q$ is irreducible if $\forall i, j \in S$, there exists an $n_0 < \infty$ such that $P_{ij}^{(n_0)} > 0$. All regular Markov chains are irreducible, but the converse is not true.

# 12 Week 12

## 12.1 Lecture 20. Mon Nov 18

**Example 12.1.** Let

$$Q = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\ 1/3 & 2/3 & 0 & 0 \\ 0 & 0 & 1/4 & 3/4 \\ 0 & 0 & 1/2 & 1/2 \end{pmatrix}.$$

The Markov chain defined by $Q$ is not irreducible.

**Exercise 12.1.** Find a Markov chain defined by a transition matrix $Q$ that is irreducible and not regular.

*Solution.* Consider

$$Q = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

We know

$$Q^2 = Q^4 = \ldots = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

and

$$Q = Q^3 = Q^5 = \ldots = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

But $P_{ii}^{(n)} > 0 \iff n$ is even. We say that this Markov chain has period 2, since $Q$ is periodic with period 2. □

### 12.1.1 Passage Time

**Definition 12.2.** The first passage time from state $i$ to state $k$, $i \neq k$, is given by

$$T_{ik} = \min\{n > 0 \mid X_n = k, X_0 = i\}.$$

The mean first passage time is

$$\mathcal{M}_{ik} = \mathbf{E}\left[T_{ik}\right].$$

**Definition 12.3.** The first recurrence time of $i$ is defined by

$$T_i = \min\{n > 0 \mid X_n = i, X_0 = i\}.$$

The mean recurrence time of $i$ is defined by

$$\mathcal{M}_i = \mathbf{E}\left[T_i\right].$$

**Exercise 12.2.** Let

$$Q = \begin{pmatrix} 1/3 & 2/3 \\ 1/4 & 3/4 \end{pmatrix}.$$

Find $\mathcal{M}_{1,2}, \mathcal{M}_1$.

*Proof.* We can quickly deduce that

$$P(T_{12} = r) = \left(\frac{1}{3}\right)^{r-1} \frac{2}{3}.$$

Hence, $T_{12} \sim \text{Geom}\left(\frac{2}{3}\right)$, so $\mathbf{E}\left[T_{12}\right] = \frac{3}{2}$. For $\mathcal{M}_1$, use the fact that

$$\sum_{n=1}^{\infty} nr^n = \frac{r}{(1-r)^2},$$

for $|r| < 1$. □

### 12.1.2 Stationary Distributions

What is the long run behavior of a Markov chain? Recall that we defined

$$\alpha_1 = P(X_0 = 1), \quad \alpha_i = P(X_0 = i), \quad i \in S,$$

and $\alpha = (\alpha_1, \ldots, \alpha_n)$ to be the distribution of $X_0$. We know also that

$$\alpha_j^{(n)} = P(X_n = i), \quad \alpha^{(n)} = \alpha Q^n.$$

**Definition 12.4.** Let $S = (1, 2, \ldots, M)$. We say that $\pi = (\pi_1, \pi_2, \ldots, \pi_M)$ is a stationary distribution of a Markov chain $X$ with a transition matrix $Q$ if

1. $\pi_i \geq 0, \quad \forall i \in S$,
2. $\sum_{i \in S} \pi_i = 1$,
3. $\pi Q = \pi$, that is, $\pi$ is an eigenvector of $Q$ corresponding to $\lambda = 1$.

Suppose then that

$$\pi = (P(X_0 = 1), P(X_0 = 2), \ldots, P(X_0 = M)),$$

and $\pi Q = \pi$. Then

$$\begin{aligned} \pi Q^2 &= \pi Q = \pi, \\ \pi Q^3 &= \pi Q^2 = \pi Q = \pi, \\ &\vdots \end{aligned}$$

and so forth.

**Example 12.5.** Find the stationary distributions of the following:

$$Q = \begin{pmatrix} 0.8 & 0.2 \\ 0.3 & 0.7 \end{pmatrix}.$$

## 12.2 Lecture 21. Wed Nov 20

**Exercise 12.3.** Let $\{X_n\}, \{Y_n\}$ be regular independent Markov chains with corresponding transition matrices $P, Q$ respectively. Prove that $Z_n = (X_n, Y_n)$ are regular.

*Proof.* Since $\{X_n\}$ is regular, there exists $n_0 \geq 0$ such that $P_{ij}^{(n_0)} > 0, \forall i, j$. Also, there exists $m_0 > 0$ such that $Q_{ij}^{(m_0)} > 0, \forall i, j$. Let $R$ be the transition matrix of $Z$. We want to prove that there exists $N$ such that $R_{ij}^{(N)} > 0$, for all $i, j$. By the Markov property,

$$\begin{aligned} P(Z_{n+1} = (i, j) \mid Z_n = (k, \ell)) &= P(X_{n+1} = i, Y_{n+1} = j \mid X_n = k, Y_n = \ell) \\ &= P(X_{n+1} = i \mid X_N = k) P(Y_{n+1} = j \mid Y_n = \ell) \\ &= P_{ki} Q_{\ell j}. \end{aligned}$$

Similarly, we can show that

$$P(Z_N = (i, j) \mid Z_0 = (k, \ell)) = P_{ki}^{(n)} Q_{\ell j}^{(n)}.$$

Then there exists $n_0, m_0 \in \mathbb{Z}^+$ such that $P_{ki}^{(n_0)} > 0, Q_{\ell j}^{(n_0)}$. We know $P_{ki}^{(n_0)}$ are elements of $P^{n_0} > 0$, and $Q_{\ell j}^{(m_0)}$ are elements of $Q^{m_0} > 0$. Also,

$$(P^{n_0})^{m_0} = P^{n_0 m_0} > 0, \quad (Q^{m_0})^{n_0} = Q^{n_0 m_0} > 0.$$

Set $N = n_0 m_0$. Then

$$\begin{aligned} P(Z_N = (i,j) \mid Z_0 = (k,\ell)) &= P_{ki}^{(N)} Q_{\ell j}^{(N)} \\ &= P_{ki}^{(n_0 m_0)} Q_{\ell j}^{(n_0 m_0)} \\ &> 0. \end{aligned}$$

□

### 12.2.1 Classification of States

**Definition 12.6.** We say state $j$ is accessible by state $i$ if there exists $n \in \mathbb{Z}^+$ such that $P_{ij}^{(n)} > 0$. (Recall that $P_{ij}^{(n)} = P(X_n = j \mid X_0 = i)$). We write $i \longrightarrow j$ in this case.

**Definition 12.7.** We say $j$ communicates with $i$ if $i \longrightarrow j$ and $j \longrightarrow i$. We write $j \longleftrightarrow j$.

**Theorem 12.8.** *A Markov chain is irreducible if and only if $i \longleftrightarrow j$ for all states $i, j$.*

**Definition 12.9.** Denote $f_i = P(X_n = i, \text{ for some } n \geq 1 \mid x_0 = i)$. We call $f_i$ the probability that the Markov chain eventually reenters state $i$, given that it starts at $i$.

**Definition 12.10.** The state $i$ is recurrent if $f_i = 1$. I.e. the Markov chain enters $i$ infinitely many times often as the number of steps we take tends towards infinity.

**Definition 12.11.** If $f_i < 1$, the state $i$ is called a transient state. In other words, the Markov chain enters state $i$ finitely many times. Write

$$I_n = \begin{cases} 1, & X_n = i \mid X_0 = i \\ 0, & X_n \neq i \mid X_0 = i. \end{cases}$$

The expected number of times state $i$ is visited is

$$\mathbf{E}\left[\sum_{n=0}^{\infty} I_n\right] = \sum_{n=0}^{\infty} \mathbf{E}\left[I_n\right] = \sum_{n=0}^{\infty} P(X_n = i \mid X_0 = i) = \sum_{n=0}^{\infty} P_{ii}^{(n)}.$$

**Theorem 12.12.** *The following hold true:*

1. *state $i$ is regular $\iff \sum_{n=0}^{\infty} P_{ii}^{(n)} = \infty$,*
2. *state $i$ is transient $\iff \sum_{n=0}^{\infty} P_{ii}^{(n)} < \infty$.*

**Theorem 12.13.** *If state $i$ and $i$ communicates with $j$, ($i \longleftrightarrow j$), then $j$ is also recurrent.*

# 13 Week 13

## 13.1 Lecture 22. Mon Dec 2

### 13.1.1 Markov Chain Recurrence Classes

Today marks the end of course material. Continuing what we started before break, let us prove the previous theorem.

**Theorem 13.1.** *If $i$ is recurrent and $i \longleftrightarrow j$, then $j$ is recurrent.*

*Proof.* The goal is to show that

$$\sum_{n=1}^{\infty} P_{jj}^{(n)} = \infty.$$

Let $i$ be recurrent, and $i \longleftrightarrow j$. Then there exists $m, k$ such that

$$P_{ij}^{(m)} > 0, \quad P_{ji}^{(k)} > 0.$$

Let $n > 0$. Consider $P_{jj}^{(m+n+k)}$. We know that

$$P_{jj}^{(m+n+k)} \geq P_{ji}^{(k)} P_{ii}^{(n)} P_{ij}^{(m)},$$

since the right hand side counts a subset of the paths we can take to attain the left hand side. Now,

$$\begin{aligned}
\sum_{n=1}^{\infty} P_{jj}^{(m+n+k)} &\geq \sum_{n=1}^{\infty} P_{ji}^{(k)} P_{ii}^{(n)} P_{ij}^{(m)} \\
&= \underbrace{P_{ji}^{(k)} P_{ij}^{(m)}}_{>0} \underbrace{\sum_{n=1}^{\infty} P_{ii}^{(n)}}_{\text{diverges, since } i \text{ is recurrent}} \\
&> \infty.
\end{aligned}$$

□

**Definition 13.2.** We say $i, j$ are in the same class if $i \longleftrightarrow j$.

**Corollary 13.2.1.** *On an irreducible Markov chain, the relation $\longleftrightarrow$ is an equivalence relation on the state space $S$.*

**Corollary 13.2.2.** *Let $C$, $D$ be two classes of an irreducible Markov chain. Then either $C = D$, or $C \cap D = \varnothing$.*

**Exercise 13.1.** Determine all of the classes of the Markov chain represented by

1.
$$P = \begin{pmatrix} 0.5 & 0.5 & 0 & 0 \\ 0.3 & 0.6 & 0.1 & 0 \\ 0 & 0 & 0.2 & 0.8 \\ 0 & 0 & 0.9 & 0.1 \end{pmatrix},$$

2.

$$P = \begin{pmatrix} 0.5 & 0.5 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 \\ 0.25 & 0.25 & 0.25 & 0.25 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

**Definition 13.3.** We say that a class is closed if

$$P_{ij} = 0, \quad \forall i \in C, \forall j \notin C.$$