

# **Course Notes**

## **Introduction to Probability**

*Alex Rutar*

BSM Fall 2018

# Contents

<b>1</b>	<b>Fundamentals</b>	<b>3</b>
1.1	Basic Principles . . . . .	3
1.1.1	Probability Spaces . . . . .	3
1.1.2	$\Omega$ . . . . .	3
1.1.3	$\mathcal{F}$ . . . . .	4
1.1.4	$\mathbb{P}$ . . . . .	4
1.1.5	Consequences . . . . .	4
1.1.6	Examples with Finite Uniform Probabilities . . . . .	6
1.2	Conditional Probability . . . . .	8
1.2.1	Basic Principles . . . . .	8
1.2.2	Bayes' Formula . . . . .	10
1.3	Independent Events . . . . .	12
1.3.1	Definitions . . . . .	12
1.3.2	Independent Trials . . . . .	12
1.3.3	Random Walks . . . . .	13
1.3.4	Conditional Independence . . . . .	14
<b>2</b>	<b>Discrete Random Variables</b>	<b>17</b>
2.1	Basics . . . . .	17
2.1.1	Expected Value . . . . .	18
2.1.2	Variance . . . . .	19
2.2	The Binomial Distribution . . . . .	20
2.2.1	Basic Properties . . . . .	20
2.2.2	Bernoulli's Law of Large Numbers . . . . .	21
2.3	Poisson Distribution . . . . .	22
2.3.1	Basic Properties . . . . .	22
2.3.2	Poisson Process . . . . .	23
2.4	Additional Discrete Distributions . . . . .	25
2.4.1	Geometric Distribution . . . . .	25
2.4.2	Negative Binomial Distribution . . . . .	26
2.4.3	Hypergeometric distribution . . . . .	26
<b>3</b>	<b>Continuous Random Variables</b>	<b>29</b>
3.1	Cumulative Distribution Function . . . . .	29
3.1.1	Properties of the CDF . . . . .	29



# Chapter 1

## Fundamentals

### 1.1 Basic Principles

#### 1.1.1 Probability Spaces

A probability space is a triple  $(\Omega, \mathcal{F}, \mathbb{P})$ .

#### 1.1.2 $\Omega$

$\Omega$  is a set, called the sample space, and  $\omega \in \Omega$  are called outcomes and  $A \subset \Omega$  are called events.

**Ex. 1.1.1** A horserace with 3 horses,  $a, b, c$ , has  $\Omega = \{(a, b, c), (a, c, b), \dots, (c, b, a)\}$ . Then  $|\Omega| = 6$  and  $A = \{a \text{ wins the race}\} = \{(a, b, c), (a, c, b)\}$ .

**Ex. 1.1.2** Roll two fair dice, a white die and a yellow die. Then  $\Omega = \{(1, 1), (1, 2), \dots, (6, 6)\}$  and  $|\Omega| = 36$ .

**Ex. 1.1.3** Continue flipping a coin until there is a head. Then

$$\Omega = \{(H), (T, H), (T, T, H), \dots\}$$

Then define

$$A = \{\text{there are an even number of rolls}\} = \{(T, H), (T, T, T, H), \dots\}$$

**Ex. 1.1.4** Consider  $\Omega = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 \leq 100\}$ . Then  $A = \{\text{you score 50 points}\} = \{(x, y) \mid x^2 + y^2 \leq 1\}$ .

**Def'n. 1.1.5** If  $A \cap B = \emptyset$ , we say that  $A$  and  $B$  are **mutually exclusive** events. If  $A \subset B$ , we say that  $A$  **implies**  $B$ .

Write  $A^c = \Omega \setminus A$ . Recall distributivity, the deMorgan relations, etc.

### 1.1.3 $\mathcal{F}$

$\mathcal{F}$  is a collection of subsets of  $\Omega$ , which denote the events that we consider.

- If  $\Omega$  is countable, then typically  $\mathcal{F}$  is just the collection of all subsets of  $\Omega$ .
- If  $\Omega$  is a domain in  $\mathbb{R}^n$ , then it is a strict subset of  $\mathbb{R}^n$ .

In any case,  $\mathcal{F}$  has to be closed under the following operations:

1.  $\Omega \in \mathcal{F}$
2. If  $A \in \mathcal{F}$ , then  $A^c \in \mathcal{F}$
3. If  $A_1, A_2, \dots \in \mathcal{F}$ , then  $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$ .

in other words, that  $\mathcal{F}$  is a  $\sigma$ -algebra.

### 1.1.4 $\mathbb{P}$

Finally,  $\mathbb{P} : \mathcal{F} \rightarrow \mathbb{R}$  is a function that satisfies 3 axioms:

1. For any  $A \in \mathcal{F}$ , then  $\mathbb{P}(A) \geq 0$
2.  $\mathbb{P}(\Omega) = 1$
3. ( $\sigma$ -additivity) Let  $A_1, A_2, A_3, \dots$  be a sequence of mutually exclusive events. Then

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$

### 1.1.5 Consequences

- $\mathbb{P}(A^c) + \mathbb{P}(A) = \mathbb{P}(A \cup A^c) = \mathbb{P}(\Omega) = 1$ .
- If  $A \subset B$ , then  $\mathbb{P}(A) \leq \mathbb{P}(B)$  since  $\mathbb{P}(B) = \mathbb{P}((A^c \cap B) \cup (A \cap B)) = \mathbb{P}(A^c \cap B) + \mathbb{P}(A \cap B) = \mathbb{P}(A^c \cap B) + \mathbb{P}(A)$
- For any  $A, B$ , we have

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

Similarly,

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

which generalizes arbitrarily:

$$\mathbb{P}\left(\bigcup_{i=1}^n A_i\right) = \sum_{r=1}^n (-1)^{r+1} \sum_{1 \leq i_1 < i_2 < \dots < i_r \leq n} \mathbb{P}(A_{i_1} \cap \dots \cap A_{i_r})$$

**PROOF** We have already proved the base case for  $n = 2$ , so assume the formula holds for a union of  $n$  events. Then

$$\mathbb{P}(A_1 \cup \dots \cup A_n \cup A_{n+1}) = \mathbb{P}(A_1 \cup \dots \cup A_n) + \mathbb{P}(A_{n+1}) - \mathbb{P}((A_1 \cup \dots \cup A_n) \cap A_{n+1})$$

We can distribute the first and third terms using the induction hypothesis, and the result follows.  $\square$

**Def'n. 1.1.6** We say  $D_1, D_2, \dots$  is a **decreasing** sequence of events of  $D_{k+1} \subset D_k$ . We say  $D_1, D_2, \dots$  is a **increasing** sequence of events of  $D_{k+1} \supset D_k$ .

Let  $\lim_{n \rightarrow \infty} D_n = \bigcap_{n=1}^{\infty} D_n$  and  $\lim_{n \rightarrow \infty} I_n = \bigcup_{n=1}^{\infty} I_n$ .

**Prop. 1.1.7**  $\sigma$ -additivity implies that for any increasing sequence,

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} I_n\right) = \lim_{n \rightarrow \infty} \mathbb{P}(I_n)$$

and similarly for any decreasing sequence

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} D_n\right) = \lim_{n \rightarrow \infty} \mathbb{P}(D_n)$$

**PROOF** Note that (2) implies (1): if  $D_k$  is a decreasing sequence, then  $I_k = D_k^c$  is an increasing sequence and

$$\left(\lim_{n \rightarrow \infty} D_n\right)^c = \left(\bigcap_{n=1}^{\infty} D_n\right)^c = \bigcup_{n=1}^{\infty} I_n = \lim_{n \rightarrow \infty} I_n$$

and taking probabilities,

$$\mathbb{P}\left(\lim_{n \rightarrow \infty} D_n\right) = 1 - \mathbb{P}\left(\lim_{n \rightarrow \infty} I_n\right) = 1 - \lim_{n \rightarrow \infty} \mathbb{P}(I_n) = \lim_{n \rightarrow \infty} \mathbb{P}(D_n)$$

To prove that  $\sigma$ -additivity implies (1), let  $I_1, I_2, \dots$  be increasing. Let  $A_1 = I_1$  and for  $k \geq 2$  let  $A_k = I_k \setminus I_{k-1}$ . Then  $A_1, A_2, \dots$  are mutually exclusive and for any  $k \geq 1$ ,

$$\bigcup_{k=1}^K A_k = I_K$$

Thus

$$\bigcup_{k=1}^{\infty} A_k = \lim_{n \rightarrow \infty} I_n$$

Now note that  $\mathbb{P}(I_K) = \sum_{k=1}^K \mathbb{P}(A_k)$  while

$$\begin{aligned} \mathbb{P}\left(\lim_{n \rightarrow \infty} I_n\right) &= \mathbb{P}\left(\bigcup_{k=1}^{\infty} A_k\right) \\ &= \sum_{k=1}^{\infty} \mathbb{P}(A_k) \\ &= \lim_{K \rightarrow \infty} \sum_{k=1}^K \mathbb{P}(A_k) \\ &= \lim_{K \rightarrow \infty} \mathbb{P}(I_K) \end{aligned}$$

$\square$

### 1.1.6 Examples with Finite Uniform Probabilities

We assume that  $\Omega = \{\omega_1, \omega_2, \dots, \omega_N\}$  and  $\mathbb{P}(\{\omega_i\}) = \mathbb{P}(\{\omega_j\})$ . Then  $\mathbb{P}(\{\omega_i\}) = \frac{1}{N}$  and  $\mathbb{P}(A) = |A|/N$ .

**Ex. 1.1.8** In an urn there are 6 blue balls and 5 red balls. Draw 3 balls out of this 11. What is the chance that among the 3 there are exactly 2 blue balls and 1 red ball?

Let us pretend that the balls are labelled, 1 through 11, and set  $\Omega$  to be all the ordered triples of disjoint elements. Then  $A = \{\text{exactly 2 blue and 1 red}\}$ , and note that  $A = A^1 \cup A^2 \cup A^3$  where  $A^i$  has a red in position  $i$  and blue in the other two positions. Now,  $|A^i| = 5 \cdot 6 \cdot 5$ , so  $|A| = 3 \cdot 6 \cdot 5 \cdot 6$  and

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{3 \cdot 6 \cdot 5 \cdot 6}{11 \cdot 10 \cdot 9}$$

We now suppose that  $\Omega = \{\Lambda \subset \{1, \dots, 11\} \mid |\Lambda| = 3\}$ , so  $|\Omega| = \binom{11}{3}$ . Now

$$A = \{\Lambda_1 \cup \Lambda_2 \mid \Lambda_1 \subset \{1, \dots, 6\}, |\Lambda_1| = 2, \Lambda_2 \subset \{7, \dots, 11\}, |\Lambda_2| = 1\}$$

So  $|A| = \binom{6}{2} \cdot 5$ .

**Ex. 1.1.9** Consider a group of  $N$  people. What is the chance that there is at least one pair among them who have the same birthday?

Define  $\Omega = \{(i_1, i_2, \dots, i_N) \mid i_j \in \{1, \dots, 365\}\}$ . We want  $A = \{\text{there is at least one common birthday}\}$ . We can write

$$A^c = \{(i_1, \dots, i_N) \in \Omega \mid i_j \neq i_k \forall j \neq k\}$$

Then  $|A^c| = 365 \cdot 364 \cdots (365 - N + 1)$  and

$$P_N = \mathbb{P}(A) = 1 - \mathbb{P}(A^c) = 1 - \frac{365 \cdot 364 \cdots (365 - N + 1)}{365^N}$$

**Ex. 1.1.10** Suppose we have  $N$  people at a party. The following day, everyone leaves one after another, and chooses a single phone from a pile. What is the chance that nobody chooses her own phone?

Define  $\Omega = \{(i_1, \dots, i_N) \mid \text{permutations of } \{1, \dots, N\}\}$ , so  $\omega = (i_1, \dots, i_k)$  means person  $k$  chooses phone  $i_k$ . Then  $|\Omega| = N!$ . Fix  $B = \{\text{nobody picks her/his phone}\}$ . Define  $A_1 = \{\text{person 1 picks his phone}\}$ , so  $|A_1| = (N - 1)!$ , and similarly for  $A_2$ , etc. Then  $B = A_1^c \cap A_2^c \cdots \cap A_N^c = (A_1 \cup \dots \cup A_N)^c$ , and  $\mathbb{P}(A_i) = \frac{1}{N}$ . Now in general,

$$\mathbb{P}(A_{i_1} \cap \dots \cap A_{i_k}) = \frac{(N - k)!}{N!}$$

for  $i_k$  distinct. Thus we now have

$$\begin{aligned} \mathbb{P}(B) &= 1 - \mathbb{P}(A_1 \cup A_2 \cup \dots \cup A_N) \\ &= 1 - \sum_{r=1}^N (-1)^{r+1} \sum_{1 \leq i_1 < i_2 < \dots < i_r \leq N} \mathbb{P}(A_{i_1} \cap \dots \cap A_{i_r}) \\ &= \sum_{r=1}^N (-1)^{r+1} \binom{N}{r} \frac{(N - r)!}{N!} \\ &= \sum_{r=1}^N (-1)^{r+1} \frac{1}{r!} \end{aligned}$$

so that

$$\mathbb{P}(B) = 1 + \sum_{r=1}^N (-1)^r \frac{1}{r!} = \sum_{r=0}^N (-1)^r \frac{1}{r!}$$

Thus  $\lim_{N \rightarrow \infty} \mathbb{P}(B) = \frac{1}{e}$ .

**Ex. 1.1.11 (Round table seating)** Consider a round table with 20 seats, and 10 married couples sit. What is the change that no couples sit together?

Define  $\Omega = \{\text{permutations of } \{1, \dots, 20\}\} / \sim$  where  $(i_1, \dots, i_{20}) \sim (i_{20}, i_1, \dots, i_{19})$ . Then  $|\Omega| = 19!$ . Define  $B = \{\text{no couples together} = A_1^c \cap A_2^c \cap \dots \cap A_{10}^c\}$ , where

$$A_k = \{\text{the } k\text{th woman sits next to her spouse}\}$$

so that

$$\mathbb{P}(B) = 1 - \mathbb{P}(A_1 \cup \dots \cup A_{10})$$

Note that

$$\mathbb{P}(A_i) = \frac{18! \cdot 2}{19!} = \frac{2}{19}$$

by “joining” the couple together, arranging them around the table, and permuting the couple internally. This generalizes to

$$\mathbb{P}(A_{i_1} \cap \dots \cap A_{i_r}) = \frac{2^r (19-r)!}{19!}$$

Then by inclusion-exclusion,

$$\mathbb{P}(B) = 1 - \binom{10}{1} \frac{18! \cdot 2}{19!} + \binom{10}{2} \frac{17! \cdot 2^2}{19!} - \binom{10}{3} \frac{16! \cdot 2^3}{19!} \dots + \binom{10}{10} \frac{9! \cdot 2^{10}}{19!} \approx 0.339$$

**Ex. 1.1.12 (Poker hand probabilities)** A poker hand is a straight if the 5 cards are of increasing value and not all of the same suit, starting with A, 2, 3, 4, ..., 10.

Define  $\Omega = \{5 \text{ element subsets of the } 52 \text{ cards}\}$ . Then  $|\Omega| = \binom{52}{5}$ . Thus

$$\mathbb{P}(\text{straight}) = \frac{10 \cdot (4^5 - 4)}{\binom{52}{5}}$$

$$\mathbb{P}(\text{full house}) = \frac{13 \cdot 12 \cdot \binom{4}{3} \cdot \binom{4}{2}}{\binom{52}{5}}$$

**Ex. 1.1.13 (Bridge hand probabilities)** In bridge, each of the 4 players get 13 cards. Let  $\Omega = \{13 \text{ cards that North gets}\}$ .

1. Let  $E$  denote the event that North receives all spades. Then

$$\mathbb{P}(E) = \frac{1}{\binom{52}{13}}$$



2. Now let  $E$  denote the event that north does not receive all 4 suits of any value, so  $E^c$  is the event that North receives all 4 of some suit. Thus  $\mathbb{P}(E) = 1 - \mathbb{P}(E^c)$ . Let  $V_k$  denote the event that North gets all suits of suit  $k$ . Then

$$\begin{aligned}\mathbb{P}(V_1) &= \frac{\binom{48}{9}}{\binom{52}{13}} \\ \mathbb{P}(V_1 \cap V_2) &= \frac{\binom{44}{5}}{\binom{52}{13}} \\ \mathbb{P}(V_1 \cap V_2 \cap V_3) &= \frac{\binom{40}{1}}{\binom{52}{13}}\end{aligned}$$

So that, by Inclusion Exclusion,

$$1 - \mathbb{P}(V_1 \cup V_2 \cup \dots \cup V_{13}) = 1 - \frac{\binom{48}{9}}{\binom{52}{13}} \cdot 13 + \binom{13}{2} \frac{\binom{44}{5}}{\binom{52}{13}} - \binom{13}{3} \frac{40}{\binom{52}{5}}$$

What is the chance that each player receives one ace? There are

$$\frac{52!}{13!13!13!13!}$$

possible hands. There are  $4!$  ways to arrange the aces, which gives

$$\mathbb{P}(E) = \frac{4! \binom{48}{12,12,12,12}}{\binom{52}{13,13,13,13}}$$

## 1.2 Conditional Probability

### 1.2.1 Basic Principles

Suppose we roll two fair dice. Then  $\mathbb{P}(\text{the sum is } 10) = \frac{3}{36} = \frac{1}{12}$ . Suppose instead that the white dice is rolled first, and it turns up 6. Now the probability that the sum is 10 is now  $1/6$ .

**Def'n. 1.2.1** Given an event  $E$  with  $\mathbb{P}(E) > 0$ , for any event  $F$ , let  $\mathbb{P}(F|E) = \frac{\mathbb{P}(F \cap E)}{\mathbb{P}(E)}$ . We call this the **conditional probability of  $F$  given  $E$** .

**Prop. 1.2.2** Fix  $E$  with  $\mathbb{P}(E) > 0$  and consider  $\mathbb{P}(\cdot|E) : \mathcal{F} \rightarrow \mathbb{R}$ . This function satisfies the axioms of probability.

**PROOF** 1.  $\mathbb{P}(F|E) \geq 0$  for all  $F \in \mathcal{F}$ .

$$2. \mathbb{P}(\Omega|E) = \frac{\mathbb{P}(E \cap \Omega)}{\mathbb{P}(E)} = 1$$

3. If  $F_1, F_2, \dots$  are mutually exclusive, then

$$\begin{aligned}\mathbb{P}\left(\bigcup_{i=1}^{\infty} F_i | E\right) &= \frac{\mathbb{P}((\bigcup_{i=1}^{\infty} F_i) \cap E)}{\mathbb{P}(E)} \\ &= \frac{\mathbb{P}(\bigcup_{i=1}^{\infty} (E \cap F_i))}{\mathbb{P}(E)} \\ &= \sum_{n=1}^{\infty} \frac{\mathbb{P}(F_n \cap E)}{\mathbb{P}(E)} \\ &= \sum_{n=1}^{\infty} \mathbb{P}(F_n | E)\end{aligned}\quad \square$$

**Prop. 1.2.3** We have  $\mathbb{P}(E \cap F) = \mathbb{P}(F|E) \cdot \mathbb{P}(E)$ , and more generally

$$\mathbb{P}(E_n \cap E_{n-1} \cap \dots \cap E_1) = \mathbb{P}(E_n | E_{n-1} \cap \dots \cap E_1) \dots \mathbb{P}(E_3 | E_2 \cap E_1) \mathbb{P}(E_2 | E_1) \mathbb{P}(E_1)$$

PROOF This follows by induction from the definition of conditional probability.  $\square$

**Ex. 1.2.4** Andrew and Bob play for the college basketball team. They get two T-shirts each, in closed bags. Any T-shirt can be black or white, with 50-50 chance. Andrew prefers black, but Bob has no preference. The following day, Andrew shows up with a black shirt on. What is the chance that Andrew's other shirt is black?

SOL'N We have  $\Omega = \{(B, B), (B, W), (W, B), (W, W)\}$  which is reduced to  $\{(B, B), (B, W), (W, B)\}$ , so the answer is  $1/3$ . To make this transparent, consider

$$\begin{aligned}A_1 &= \{\text{Andrew has at least one black shirt}\} \\ A_2 &= \{\text{Both of Andrew's shirts are black}\} \\ A_3 &= \{\text{Andrew has a black shirt on}\}\end{aligned}$$

so in Andrew's case,  $A_1 = A_3$  and  $\mathbb{P}(A_2 | A_3) = \mathbb{P}(A_2 | A_1)$ .

**Ex. 1.2.5 (Polya's Urn)** Initially, we have two balls, 1 red, 1 blue, in the urn. For the first draw, pick one, check its color, and put it back and put another ball of the same color into the urn. What is the probability that the first three balls are red, blue, red (in this order)?

SOL'N Let  $R_i, B_i$  denote the  $i^{th}$  draw is red or blue respectively. Then

$$\mathbb{P}(R_3 \cap B_2 \cap R_1) = \mathbb{P}(R_3 | B_2 \cap R_1) \mathbb{P}(B_2 | R_1) \mathbb{P}(R_1) = \frac{1}{2} \frac{1}{3} \frac{1}{2} = \frac{1}{12}$$

**Ex. 1.2.6** What is the probability in bridge, each of the players gets one ace?

SOL'N Let  $E_4$  denote the event in which every player gets an ace. Then

$$\begin{aligned} & E_4 \\ & \cap \\ & E_3 = \{\text{Aces of spaces, hearts, and diamonds are at 3 different players.}\} \\ & \cap \\ & E_2 = \{\text{Aces of spaces, hearts, and diamonds are at 2 different players.}\} \\ & \cap \\ & E_1 = \Omega \end{aligned}$$

so that  $\mathbb{P}(E_4) = \mathbb{P}(E_4 \cap E_3 \cap E_2 \cap E_1) = \mathbb{P}(E_4|E_3)\mathbb{P}(E_3|E_2)\mathbb{P}(E_2|E_1)\mathbb{P}(E_1)$ .

### 1.2.2 Bayes' Formula

**Ex. 1.2.7** Consider an insurance company, which classifies people into accident prone drivers (30%) and non-accident-prone drivers, (70%). For accident prone drivers, the chance of being involved in an accident within a year is 0.2, while for non-accident-prone drivers, the chance of being involved in an accident is 0.1. Now suppose we have a new policyholder.

1. What is the probability that the policyholder is involved in an accident within a year?
2. The policyholder was involved in an accident?

SOL'N 1.  $B = \{\text{accident in 2018}\}$ ,  $A = \{\text{the policyholder is accident prone}\}$ . Then

$$\mathbb{P}(B) = \mathbb{P}(B \cap A) + \mathbb{P}(B \cap A^c) = \mathbb{P}(B|A)\mathbb{P}(A) + \mathbb{P}(B|A^c)\mathbb{P}(A^c) = 0.2 \cdot 0.3 + 0.1 \cdot 0.7 = 0.13$$

2. Now

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B|A)\mathbb{P}(A) + \mathbb{P}(B|A^c)\mathbb{P}(A^c)} = \frac{0.2 \cdot 0.3}{0.13} = \frac{6}{13}$$

**Prop. 1.2.8** Suppose  $A_1, A_2, \dots, A_n \in \mathcal{F}$  form a partition of  $\Omega$ . Given such a partition, for any  $B \in \mathcal{F}$ ,

$$\mathbb{P}(B) = \sum_{i=1}^n \mathbb{P}(B \cap A_i) = \sum_{i=1}^n \mathbb{P}(B|A_i) \cdot \mathbb{P}(A_i)$$

Then for any  $k \in [n]$ ,

$$\mathbb{P}(A_k|B) = \frac{\mathbb{P}(B \cap A_k)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B|A_k) \cdot \mathbb{P}(A_k)}{\sum_{i=1}^n \mathbb{P}(B|A_i) \cdot \mathbb{P}(A_i)}$$

**Ex. 1.2.9** Roll a fair dice. There is a urn with one white ball in it. If the die turns up 1, 3, or 5, put one black ball into the urn. If it turns up 2 or 4, put 3 black and 5 white, and if it turns up 6, put 5 black and 5 white.

SOL'N Write

$$\begin{aligned} A_1 &= \{1, 3 \text{ or } 5 \text{ rolled}\} \\ A_2 &= \{2 \text{ or } 4 \text{ rolled}\} \\ A_3 &= \{6 \text{ rolled}\} \\ B &= \{\text{black ball rolled}\} \end{aligned}$$

so that

$$\begin{aligned}\mathbb{P}(A_3|B) &= \frac{\mathbb{P}(B|A_3)\mathbb{P}(A_3)}{\mathbb{P}(B|A_1) \cdot \mathbb{P}(A_1) + \mathbb{P}(B|A_2) \cdot \mathbb{P}(A_2) + \mathbb{P}(B|A_3) \cdot \mathbb{P}(A_3)} \\ &= \frac{5/6 \cdot 1/6}{1/2 \cdot 1/2 + 3/4 \cdot 1/3 + 5/6 \cdot 1/6} \\ &= \frac{5}{23}\end{aligned}$$

**Ex. 1.2.10** There is a blood test for a rare but serious disease. Only 1/10000 people have this disease. Suppose the test is 100% effective, so if someone is tested ill, it is positive with 100% chance. Suppose there is also a 1% chance of false positive. A new patient is tested, and tests positive. What are the odds that she has the disease?

Sol'N Let  $A = \{\text{the person is ill}\}$  and  $B = \{\text{the test is positive}\}$ . Then

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B|A)\mathbb{P}(A) + \mathbb{P}(B|A^c)\mathbb{P}(A^c)} = \frac{1 \cdot 0.0001}{1 \cdot 0.0001 + 0.01 \cdot 0.9999}$$

**Ex. 1.2.11 (Monty Hall paradox)** There are three doors: one of them hides a prize, and two hide nothing. Pick a door. The announcer then reveals another door not containing a prize. Is it better to stay or switch?

Sol'N Write  $A_i = \{\text{door } i \text{ hides the prize}\}$ , and  $B_2 = \{\text{door 2 is opened}\}$ . Then

$$\begin{aligned}\mathbb{P}(A_1|B_2) &= \frac{\mathbb{P}(B_2|A_1)\mathbb{P}(A_1)}{\mathbb{P}(B_2|A_1)\mathbb{P}(A_1) + \mathbb{P}(B_2|A_2)\mathbb{P}(A_2) + \mathbb{P}(B_2|A_3)\mathbb{P}(A_3)} \\ &= \frac{1/2 \cdot 1/3}{1/2 \cdot 1/3 + 0 + 1 \cdot 1/3} = \frac{1}{3}\end{aligned}$$

but

$$\begin{aligned}\mathbb{P}(A_3|B_2) &= \frac{\mathbb{P}(B_2|A_3)\mathbb{P}(A_3)}{\mathbb{P}(B_2|A_1)\mathbb{P}(A_1) + \mathbb{P}(B_2|A_2)\mathbb{P}(A_2) + \mathbb{P}(B_2|A_3)\mathbb{P}(A_3)} \\ &= \frac{1 \cdot 1/3}{1/2 \cdot 1/3 + 0 + 1 \cdot 1/3} = \frac{2}{3}\end{aligned}$$

so it is better to switch!

**Ex. 1.2.12** There is an inspection, which is 60% sure of the guilt of a certain suspect. The suspect is left-handed. There is new evidence: the criminal is left handed. Say 20% of the population is left handed; how certain should the inspector now be?

Sol'N Write  $C = \{\text{the suspect is the criminal}\}$  and  $C^c = \{\text{the criminal is someone else}\}$ . Then  $\mathbb{P}(C) = 0.6$  and  $\mathbb{P}(C^c) = 0.4$ . Let  $L = \{\text{the criminal is left-handed}\}$ . Then

$$\mathbb{P}(C|L) = \frac{\mathbb{P}(L|C)\mathbb{P}(C)}{\mathbb{P}(L)} \quad \mathbb{P}(C^c|L) = \frac{\mathbb{P}(L|C^c)\mathbb{P}(C^c)}{\mathbb{P}(L)}$$

Here, we can compute the “odds”:

$$\frac{\mathbb{P}(C|L)}{\mathbb{P}(C^c|L)} = \frac{\mathbb{P}(L|C)\mathbb{P}(C)}{\mathbb{P}(L|C^c)\mathbb{P}(C^c)}$$

Now  $\mathbb{P}(L|C) = 1$ , but  $\mathbb{P}(L|C^c) = \mathbb{P}(L) = 0.2$ , since the probability is taken a priori. Now a priori, the odds are given by  $\mathbb{P}(C)/\mathbb{P}(C^c) = 0.6/0.4$ , scaled by the factor  $\mathbb{P}(L|C)/\mathbb{P}(L|C^c) = 5$  given updated information. Thus  $\mathbb{P}(C|L) = 15/17$ .

## 1.3 Independent Events

### 1.3.1 Definitions

**Def'n. 1.3.1** The events  $A$  and  $B$  are **independent** if  $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ .

**Ex. 1.3.2** Draw a card from a deck of 52. Let

$$A = \{\text{it is a spade}\}, \quad B = \{\text{it is an ace}\}, \quad C = \{\text{it is a heart}\}$$

We have

$$\mathbb{P}(A) = \frac{1}{4}, \quad \mathbb{P}(B) = \frac{1}{13}, \quad \mathbb{P}(A \cap B) = \frac{1}{52}$$

so  $A$  and  $B$  are independent. Similarly,  $B$  and  $C$  are independent. However,  $\mathbb{P}(A \cap C) = 0 \neq 1/4$  so  $A$  and  $C$  are not independent.

**Rmk. 1.3.3** Exclusive events are quite different than independence: in fact, they are (in a sense) the opposite. Let  $\mathbb{P}(A) > 0$ . Then  $A$  and  $B$  are independent iff  $\mathbb{P}(B|A) = \mathbb{P}(B)$ . Similarly,  $A$  and  $B$  are exclusive iff  $\mathbb{P}(B|A) = 0$ .

**Ex. 1.3.4** Roll two fair dice, the yellow and the white die. Then

$$\begin{aligned} A &= \{\text{the sum is 7}\} \\ B &= \{\text{the sum is 10}\} \\ C &= \{\text{the yellow die turns up 6}\} \\ D &= \{\text{the white die turns up 6}\} \end{aligned}$$

We have  $\mathbb{P}(A) = 1/6$ ,  $\mathbb{P}(C) = 1/6$ . Then  $\mathbb{P}(A \cap C) = 1/36 = 1/6 \cdot 1/6$  so  $A$  and  $C$  are independent. Similarly,  $C$  and  $D$  are independent and  $A$  and  $D$  are independent. Thus  $A, C, D$  are pairwise independent but not independent as a triple.

**Def'n. 1.3.5** The events  $A_1, A_2, \dots$  are **independent (as a collection)** if, for any choice of indices  $1 \leq i_1 < i_2 < \dots < i_k \leq n$ , then

$$\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) = \mathbb{P}(A_{i_1})\mathbb{P}(A_{i_2}) \dots \mathbb{P}(A_{i_k})$$

### 1.3.2 Independent Trials

We have two parameters:  $n \geq 1$ , which is the number of trials, and  $p \in (0, 1)$ , which is the chance of success for an individual trial. Then  $A_k = \{\text{the } k^{\text{th}} \text{ trial is a success}\}$  so that  $\mathbb{P}(A_k) = p$  and the events  $A_1, \dots, A_n$  are independent. Our framework is to consider the space  $\Omega \times \Omega \times \dots \times \Omega$ .

**Ex. 1.3.6** Roll a fair die 10 times. Then  $A_k = \{\text{the } k^{\text{th}} \text{ roll is a 6}\}$ . Then we have

- $\mathbb{P}(\text{all } n \text{ trials are successful}) = \mathbb{P}(A_1 \cap \dots \cap A_n) = p^n$
- $\mathbb{P}(\text{there is at least one success}) = 1 - (1 - p)^n$
- $\mathbb{P}(\text{there are exactly } k \text{ success out of } n \text{ trials}) = \binom{n}{k} p^k (1 - p)^{n-k}$

Consider the case now where  $n$  is countable (infinite number of trials). Let  $S = \{\text{all trials are successful}\}$  define and  $S_n = \{\text{the first } n \text{ trials are successful}\}$ . Then  $S = \bigcap_{n=1}^{\infty} S_n$  so

$$\mathbb{P}(S) = \lim_{n \rightarrow \infty} \mathbb{P}(S_n) = \lim_{n \rightarrow \infty} p^n = 0$$

**Ex. 1.3.7** Repeatedly roll two fair dice until the sum is either 5 or 7. What is the probability that the sum is 5 when we stop?

Let  $A_i = \{\text{rolls less than } i \text{ are not 5 or 7, roll } i \text{ is 5}\}$ . Since  $\mathbb{P}(\text{roll is 5 or 7}) = 1/6 + 1/9$ , we have  $\mathbb{P}(\text{roll is not}) = 13/18$ . Thus

$$\mathbb{P}(A_i) = \left(\frac{13}{18}\right)^{i-1} \frac{5}{18}$$

so that

$$\mathbb{P}(A) = \frac{1}{9} \sum_{i=0}^{\infty} \left(\frac{13}{18}\right)^i = \frac{1}{9} \frac{1}{1 - \frac{13}{18}} = \frac{2}{5}$$

We have an alternate solution: note that  $A_1, B_1, C_1$  partition the sample space. By the law of total probability,

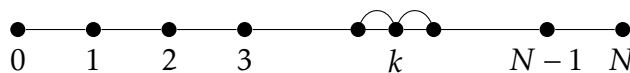
$$\begin{aligned} \mathbb{P}(D) &= \mathbb{P}(D|A_1)\mathbb{P}(A_1) + \mathbb{P}(D|A_2)\mathbb{P}(A_2) + \mathbb{P}(D|C_1)\mathbb{P}(C_1) \\ &= \mathbb{P}(B_1) + \mathbb{P}(C_1)\mathbb{P}(D) \end{aligned}$$

so that

$$\mathbb{P}(D) = \frac{\mathbb{P}(B_1)}{1 - \mathbb{P}(C_1)} = \frac{\mathbb{P}(B_1)}{\mathbb{P}(A_1) + \mathbb{P}(B_1)}$$

### 1.3.3 Random Walks

We first see the gambling interpretation. Suppose we have two players,  $A$  has initial capital  $k$  and  $B$  has initial capital  $N - k$ . At each round, a coin is flipped. If it is a head, then  $B$  gives  $A$  1 dollar, and if it is a tail,  $A$  gives  $B$  1 dollar. Repeat this until someone runs out of money.



Let  $\mathbb{P}_k^{(N)} = \mathbb{P}(\text{when starting at position } k, \text{ the probability that eventually } A \text{ wins})$ . We have  $P_0 = 0, P_N = 1$ . Then for  $1 \leq k \leq N - 1$ , we have

$$P_k = \mathbb{P}\{\text{ending at } N \text{ when starting at } k | \text{first flip is H}\} \cdot \frac{1}{2} + \mathbb{P}\{\text{end at } N \text{ if start at } k | \text{first flip is T}\} \cdot \frac{1}{2}$$

which can be written

$$\mathbb{P}_k = P_{k+1} \frac{1}{2} + P_{k-1} \frac{1}{2} \Rightarrow \frac{1}{2} (P_k - P_{k-1}) = \frac{1}{2} (P_{k+1} - P_k)$$

so, for any  $1 \leq k \leq N$ ,  $P_k - P_{k-1} = d$  and

$$1 = P_N - P_0 = P_n - P_{N-1} + P_{N-1} - P_{N-2} + \cdots + (P_1 - P_0) = N \cdot d$$

so  $d = 1/N$  and

$$P_k = P_k - P_0 = \sum_{j=1}^k (P_j - P_{j-1}) = kd = \frac{k}{N}$$

### 1.3.4 Conditional Independence

**Def'n. 1.3.8** Given  $A$  with  $\mathbb{P}(A) > 0$ , two events  $B_1$  and  $B_2$  are **conditionally independent** given  $A$  if

$$\mathbb{P}(B_1 \cap B_2 | A) = \mathbb{P}(B_1 | A) \cdot \mathbb{P}(B_2 | A)$$

**Ex. 1.3.9** 1. We have a medical test for a rare disease, and  $A = \{\text{the patient is sick}\}$  has  $\mathbb{P}(A) = 0.005$  so  $\mathbb{P}(A^c) = 0.995$ . Let  $B_1 = \{\text{the first test is positive}\}$ , so  $\mathbb{P}(B_1 | A) = 0.95$  and  $\mathbb{P}(B_1 | A^c) = 0.01$ . Then  $\mathbb{P}(A | B) \approx 0.33$ . But now let  $B_2 = \{\text{the second test is positive}\}$ . Now what is  $\mathbb{P}(A | B_1 \cap B_2)$ ? Here, the events  $B_1$  and  $B_2$  are not independent, but they are conditionally independent given either  $A$  or  $A^c$ . Thus

$$\begin{aligned} \mathbb{P}(A | B_1 \cap B_2) &= \frac{\mathbb{P}(B_1 \cap B_2 | A) \mathbb{P}(A)}{\mathbb{P}(B_1 \cap B_2)} \\ &= \frac{\mathbb{P}(B_1 | A) \mathbb{P}(B_2 | A) \mathbb{P}(A)}{\mathbb{P}(B_1 | A) \mathbb{P}(B_2 | A) \mathbb{P}(A) + \mathbb{P}(B_1 | A^c) \mathbb{P}(B_2 | A^c) \mathbb{P}(A^c)} \\ &= \frac{(0.95)^2 \cdot 0.005}{(0.95)^2 \cdot 0.005 + (0.01)^2 \cdot 0.995} \\ &\approx 0.98 \end{aligned}$$

2. Suppose

$$\begin{aligned} A &= \{\text{accident prone}\} & \mathbb{P}(A) &= 0.3 \\ A &= \{\text{not accident prone}\} & \mathbb{P}(A^c) &= 0.7 \end{aligned}$$

and let  $B_Y = \{\text{accident in year } Y\}$ . We have seen that  $\mathbb{P}(B_{2018} | A) = 0.2$  and  $\mathbb{P}(B_{2018} | A^c) = 0.1$  so  $\mathbb{P}(B_{2018}) = 0.13$ . Now

$$\begin{aligned} \mathbb{P}(B_{2019} | B_{2018}) &= \frac{\mathbb{P}(B_{2018} \cap B_{2019})}{\mathbb{P}(B_{2018})} \\ &= \frac{\mathbb{P}(B_{2019} | A) \mathbb{P}(B_{2018} | A) \mathbb{P}(A) + \mathbb{P}(B_{2019} | A^c) \mathbb{P}(B_{2018} | A^c) \mathbb{P}(A^c)}{\mathbb{P}(B_{2018} | A) \mathbb{P}(A) + \mathbb{P}(B_{2018} | A^c) \mathbb{P}(A^c)} \\ &= \mathbb{P}(B_{2019} | A) \cdot \mathbb{P}(A | B_{2018}) + \mathbb{P}(B_{2019} | A^c) \mathbb{P}(A^c | B_{2018}) \\ &= 0.2 \cdot \frac{6}{13} + 0.1 \cdot \frac{7}{13} \\ &\approx 0.15 \end{aligned}$$

**Ex. 1.3.10 (Laplace's Rule of Succession)** Suppose we have  $k + 1$  coins in a box, and coin  $i$  turns up Heads with  $\frac{i}{k}$  chance, and Tails with  $\frac{k-i}{k}$  chance (for  $i = 0, \dots, k$ ). Pick one coin, and flip the coin  $n$  times. Assume it turned Heads every  $n$  times. What is the probability that it turns up  $H$  on the  $(n + 1)^{\text{st}}$  flip?

**Sol'n** Let  $H_j$  denote the event that the  $j^{\text{th}}$  flip is  $H$  for  $j = 1, 2, \dots, n, n + 1$ . Let  $C_i$  denote the event in which the  $i^{\text{th}}$  coin is initially picked for each  $i = 1, \dots, k$ . The events  $H_j$  are

not independent, but they are conditionally independent given any of the  $C_i$ . Moreover,  $\mathbb{P}(H_j|C_k) = \frac{i}{k}$ . We thus have

$$\begin{aligned}\mathbb{P}(H_{n+1}|H_1 \cap H_2 \cap \dots \cap H_n) &= \frac{\mathbb{P}(H_1 \cap H_2 \cap \dots \cap H_{n+1})}{\mathbb{P}(H_1 \cap \dots \cap H_n)} \\ &= \frac{\sum_{i=0}^k \mathbb{P}\left(\bigcap_{j=1}^{n+1} H_j | C_i\right) \mathbb{P}(C_i)}{\sum_{i=0}^k \mathbb{P}\left(\bigcap_{j=1}^n H_j | C_i\right) \mathbb{P}(C_i)} \\ &= \frac{\sum_{i=0}^k \prod_{j=1}^{n+1} \mathbb{P}(H_j | C_i) \mathbb{P}(C_i)}{\sum_{i=0}^k \prod_{j=1}^n \mathbb{P}(H_j | C_i) \mathbb{P}(C_i)} \\ &= \frac{\sum_{i=0}^k \left(\frac{i}{k}\right)^{n+1} \frac{1}{k+1}}{\sum_{i=0}^k \left(\frac{i}{k}\right)^n \frac{1}{k+1}} \\ &:= p(k, n)\end{aligned}$$

Both the numerator and denominator of  $p(k, n)$  are sums of the form  $\sum_{i=0}^k f(i/k) \cdot 1/k$ . These are riemann sums of the function  $f$  on the interval  $[0, 1]$ , so as  $k$  goes to infinity,

$$\lim_{k \rightarrow \infty} p(k, n) = \frac{\int_0^1 x^{n+1} dx}{\int_0^1 x^n dx} = \frac{\frac{1}{n+2}}{\frac{1}{n+1}} = \frac{n+1}{n+2}$$

**Ex. 1.3.11 (Best prize problem)** Suppose we have  $N$  items, each with a distinct real value. Observe them sequentially. After observing a prize, you can take the prize, or can abandon it (and never access it again). How can you maximize the odds that you get the best prize?

**SOL'N** We will solve this over a fixed type of strategy. Define a  $k$ -strategy for each  $k = 1, \dots, N$ , in which we observe the first  $k$  items, and pick the first of the remaining ones that is better than the first  $k$ . Let  $B_k$  denote the event that we choose the best item with a  $k$ -strategy, and fix  $P_k^{(N)} = \mathbb{P}(B_k)$ . As well, let  $A_i$  denote the event that the best prize is at the  $i^{\text{th}}$  position, so  $\mathbb{P}(A_i) = 1/N$ .

Now,  $\mathbb{P}(B_k|A_j) = 0$  for  $j \leq k$ , and  $\mathbb{P}(B_k|A_j) = \frac{k}{j+1}$  for  $j > k$ . Then

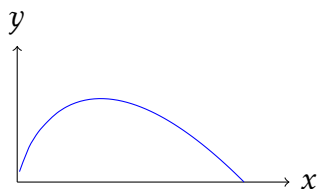
$$\begin{aligned}\mathbb{P}(B_k) &= \sum_{i=1}^N \mathbb{P}(B_k|A_i) \mathbb{P}(A_i) \\ &= \sum_{i=k}^{N-1} \frac{k}{i} \cdot \frac{1}{N} \\ &:= P_k^{(N)}\end{aligned}$$



To determine the optimal strategy, we want to choose the proportion  $x$  of items to observe. To be precise, let  $N \rightarrow \infty$  and  $k$  be such that  $k/N \rightarrow x$ . Then

$$\begin{aligned} \lim_{k/N \rightarrow x} P_k^{(N)} &= \lim_{k/N \rightarrow x} \sum_{i=k}^{N-1} \frac{k/N}{i/N} \cdot \frac{1}{N} \\ &= \lim_{k/N \rightarrow x} x \sum_{i=k}^{N-1} \frac{1}{i/N} \frac{1}{N} \\ &= x \int_x^1 \frac{1}{y} dy \\ &= -x \ln x := g(x) \end{aligned}$$

Then  $g'(x) = -\ln x - 1$  and  $g''(x) = -\frac{1}{x}$ . Then  $g'(x) = 0 \Rightarrow \ln x = -1$  so  $x = 1/e$  is a maximum since  $g''(1/e) < 0$ .



# Chapter 2

## Discrete Random Variables

### 2.1 Basics

**Def'n. 2.1.1** A *random variable* is a (measurable) function  $X : \Omega \rightarrow \mathbb{R}$ .

For example, fix  $a < b \in \mathbb{R}$  and consider the set  $\{w \in \Omega \mid \mathbb{X}(w) \in [a, b]\} \in \mathcal{F}$ .

**Ex. 2.1.2** 1. Flip three fair coins. Let  $Y$  denote the number of Heads. Then  $Y : \Omega \rightarrow \{0, 1, 2, 3\}$ .

2. Repeatedly roll a fair die until a 6 occurs. Let  $Z$  denote the number of rolls necessary. Now  $Z : \Omega \rightarrow \mathbb{N}$ .

**Def'n. 2.1.3** A random variable is **discrete** if its range is countable.

For a discrete random variable, the **probability mass function** is  $p : \mathbb{R} \rightarrow \mathbb{R}$  defined by

$$p(x) = \begin{cases} 0 & \text{if } x \text{ is not taken by } X \\ \mathbb{P}(X = x_i) & x = x_i \text{ is taken by } X \end{cases}$$

In the example  $\mathbb{P}(Y = 0) = \frac{1}{8}$ ,  $\mathbb{P}(Y = 1) = \frac{3}{8}$ ,  $\mathbb{P}(Y = 2) = \frac{3}{8}$ ,  $\mathbb{P}(Y = 3) = \frac{1}{8}$ . Note that  $\sum_{i=1}^{\infty} p(x_i) = 1$ .

In the dice example,  $\mathbb{P}(Z = k) = \left(\frac{5}{6}\right)^{k-1} \frac{1}{6}$  and indeed the geometric series sums to 1.

**Ex. 2.1.4** Each item can be one of  $N$  different types, with  $1/N$  chance independently of other items. We wish to collect all types. Let  $X$  denote the number of items needed to collect all types. We wish to determine the mass function for  $X$ .

We wish to find  $\mathbb{P}(X > n)$  for all  $n$ . Then  $\mathbb{P}(X = n) = \mathbb{P}(X > n - 1) - \mathbb{P}(X > n)$ . Now  $\{X > n\} = A_1^{(n)} \cup \dots \cup A_k^{(n)}$  where  $A_k^{(n)}$  is the event that type  $k$  has not been collected in  $n$  items.

Now

$$\begin{aligned}\mathbb{P}(X > n) &= \mathbb{P}(A_1 \cup A_2 \cup \dots \cup A_N) \\ &= \sum_{r=1}^n (-1)^{r+1} \binom{N}{r} \mathbb{P}(A_1 \cap A_2 \cap \dots \cap A_r) \\ &= \sum_{r=1}^n (-1)^{r+1} \binom{N}{r} \frac{(N-r)^n}{N^n}\end{aligned}$$

### 2.1.1 Expected Value

**Def'n. 2.1.5** The *expected value* of a discrete random variable  $X$  is given by  $\mathbb{E}(X) = \sum_{k=1}^{\infty} x_k \mathbb{P}(X = x_k)$ .

**Ex. 2.1.6** Consider two games:

1. Flip a fair coin, if H get \$100 and if T, lose
2. Roll a fair die, if 6 get \$x, otherwise, go home.

Let  $X$  denote the gain if the order is AB. We have

$$\mathbb{P}(X = 0) = \frac{1}{2}, \quad \mathbb{P}(X = 100) = \frac{1}{2} \cdot \frac{5}{6}, \quad \mathbb{P}(X = 100 + x) = \frac{1}{2} \cdot \frac{1}{6}$$

Let  $Y$  denote the gain if the order is BA. We have

$$\mathbb{P}(Y = 0) = \frac{5}{6}, \quad \mathbb{P}(Y = x) = \frac{1}{6} \cdot \frac{1}{2}, \quad \mathbb{P}(Y = 100 + x) = \frac{1}{2} \cdot \frac{1}{6}$$

so

$$\mathbb{E}(X) = 0 \cdot \frac{1}{2} + 100 \cdot \frac{5}{12} + (100 + x) \cdot \frac{1}{12} > \mathbb{E}(Y) = 0 \cdot \frac{5}{6} + x \cdot \frac{1}{12} + (x + 100) \cdot \frac{1}{12}$$

which reduces to  $500 > x$ .

**Ex. 2.1.7** Note that  $\mathbb{E}(X) = \sum_{k=1}^{\infty} x_k \mathbb{P}(X = x_k)$  if the series is absolutely convergent. For example, define  $\mathbb{P}(X = k) = \frac{1}{k(k+1)}$ , which sums to 1, but

$$\mathbb{E}(X) = \sum_{k=1}^{\infty} \frac{1}{k+1}$$

is infinite. But now, consider  $Y$  with  $\mathbb{P}(Y = 0) = 1/3$ .

**Prop. 2.1.8**  $\mathbb{E}(g(X)) = \sum_{k=1}^{\infty} g(x_k) \mathbb{P}(X = x_k)$

**PROOF** Let  $x_1, x_2, \dots$  denote the possible values of  $X$ , and  $y_1, y_2, \dots$  denote the possible values of  $Y$ . Then

$$\begin{aligned} \sum_{k=1}^{\infty} g(x_k) \mathbb{P}(X = x_k) &= \sum_{l=1}^{\infty} \sum_{x_k: g(x_k)=y_l} g(x_k) \mathbb{P}(X = x_k) \\ &= \sum_{l=1}^{\infty} y_l \sum_{x_k: g(x_k)=y_l} \mathbb{P}(X = x_k) \\ &= \sum_{l=1}^{\infty} y_l \mathbb{P}(Y = y_l) = \mathbb{E}(Y) \quad \square \end{aligned}$$

**Prop. 2.1.9**  $\mathbb{E}(aX + b) = a \mathbb{E}(X) + \mathbb{E}(b)$

**PROOF** Follows from linearity of the sum.  $\square$

## 2.1.2 Variance

Consider two random variables defined by  $\mathbb{P}(X = 1) = 1/2$  and  $\mathbb{P}(X = -1) = 1/2$  vs  $\mathbb{P}(X = 100) = 1/2$  and  $\mathbb{P}(X = -100) = 1/2$ . They both have expected value 0, so we want a value to measure the typical amount of fluctuation about the expected value. Let  $X$  be a random variable and  $\mu = \mathbb{E}(X)$ .

**Def'n. 2.1.10** We define the **variance** as  $\text{Var}(X) = \mathbb{E}[(X - \mu)^2]$ .

Note that  $(X - \mu)^2 = X^2 - 2\mu X + \mu^2$ . Then

$$\begin{aligned} \text{Var } X &= \mathbb{E}((X - \mu)^2) \\ &= \sum_{k=1}^{\infty} (x_k - 2\mu x_k + \mu^2) \mathbb{P}(X = x_k) \\ &= \sum_{k=1}^{\infty} x_k^2 \mathbb{P}(X = x_k) - 2\mu \sum_{k=1}^{\infty} x_k \mathbb{P}(X = x_k) + \mu^2 \sum_{k=1}^{\infty} \mathbb{P}(X = x_k) \\ &= \mathbb{E}(X^2) - (\mathbb{E}(X))^2 \end{aligned}$$

**Prop. 2.1.11**  $\text{Var}(aX + b) = a^2 \text{Var } X$ .

**Ex. 2.1.12** 1. Roll a fair die, so  $X$  can take  $1, 2, \dots, 6$  each with probability  $1/6$ . Then

$$\begin{aligned} \mathbb{E}(X) &= 1 \cdot \frac{1}{6} + \dots + 6 \cdot \frac{1}{6} = \frac{7}{2} \\ \mathbb{E}(X^2) &= \frac{1}{6}(1^2 + 2^2 + \dots + 6^2) = \frac{7 \cdot 13}{6} \\ \text{Var}(X) &= \frac{35}{12} \end{aligned}$$

2. Consider  $\eta = 1$  with chance  $p$  and  $0$  with chance  $1 - p$ . Then  $\mathbb{E}(\eta) = p$  and  $\mathbb{E}(\eta^2) = p$ , so  $\text{Var}(\eta) = p - p^2 = pq$ .

## 2.2 The Binomial Distribution

### 2.2.1 Basic Properties

**Def'n. 2.2.1** *The Binomial distribution has parameters  $n \geq 1$  and  $p \in (0, 1)$ . Then  $X \sim \text{Binom}(n, p)$  if  $\mathbb{P}(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$ .*

**Ex. 2.2.2** Consider a test consisting of 20 yes-no questions; you fail if you have 17 or less correct answers.

- You know the correct answer with probability 5/7
- You have the incorrect answer with probability 1/7
- You guess with probability 1/7.

On a single question, the probability that you are correct is  $5/7 + 1/14 = 11/14$ . Now

$$\begin{aligned} \mathbb{P}(\text{fail}) &= \mathbb{P}(X \leq 17) = 1 - \mathbb{P}(X = 20) - \mathbb{P}(X = 19) - \mathbb{P}(X = 18) \\ &= 1 - \binom{20}{20} \left(\frac{11}{14}\right)^{20} - \binom{20}{19} \left(\frac{11}{14}\right)^{19} \left(\frac{3}{14}\right) - \binom{20}{18} \left(\frac{11}{14}\right)^{18} \left(\frac{3}{14}\right)^2 \\ &\approx 0.8345 \end{aligned}$$

Suppose  $X \sim \text{Binom}(n, p)$ . Then

$$\begin{aligned} \mathbb{E}(X) &= \sum_{k=0}^n k \binom{n}{k} p^k (1 - p)^{n-k} \\ &= \sum_{k=1}^n \binom{n}{k} \frac{d}{dt} \Big|_{t=1} (t^k) p^k (1 - p)^{n-k} \\ &= \frac{d}{dt} \Big|_{t=1} \left( \sum_{k=0}^n \binom{n}{k} (tp)^k (1 - p)^{n-k} \right) \\ &= \frac{d}{dt} \Big|_{t=1} ((tp + 1 - p)^n) \\ &= (n(tp - p + 1)^{n-1} \cdot p) \Big|_{t=1} = np \end{aligned}$$

We can also compute

$$\begin{aligned}
 \mathbb{E}[X(X-1)] &= \sum_{k=1}^n k(k-1) \binom{n}{k} p^k (1-p)^{n-k} \\
 &= \sum_{k=2}^n \frac{d^2}{dt^2} \bigg|_{t=1} (t^k) \binom{n}{k} p^k (1-p)^{n-k} \\
 &= \frac{d^2}{dt^2} \bigg|_{t=1} \left( \sum_{k=0}^n \binom{n}{k} (tp)^k (1-p)^{n-k} \right) \\
 &= \frac{d^2}{dt^2} \bigg|_{t=1} (tp + (1-p))^n \\
 &= n(n-1)(tp + (1-p))^{n-2} p^2 \\
 &= n(n-1)p^2 \\
 &= \mathbb{E}(X^2) - \mathbb{E}(X) \\
 &= \mathbb{E}(X^2) - np
 \end{aligned}$$

so that

$$\mathbb{E}(X^2) = np^2(n-1) + np$$

and

$$\text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = n^2 p^2 - np^2 + np - n^2 p^2 = np(1-p)$$

We thus say that  $X = \eta_1 + \eta_2 + \dots + \eta_n$  where  $\eta_i = \begin{cases} 1 & \text{if trial } i \text{ is a success} \\ 0 & \text{if trial } i \text{ fail} \end{cases}$ . Thus the standard deviation scales with order  $\sqrt{n}$ .

Now let  $X \sim \text{Binom}(n, p)$ , so that

$$\frac{\mathbb{P}(X = k)}{\mathbb{P}(X = k-1)} = \frac{\binom{n}{k} p^k (1-p)^{n-k}}{\binom{n}{k-1} p^{k-1} (1-p)^{n-k+1}} = \frac{(n-k+1)p}{k(1-p)}$$

so that

$$\begin{aligned}
 1 < \frac{\mathbb{P}(X = k)}{\mathbb{P}(X = k-1)} &\Leftrightarrow k(1-p) < (n-k+1)p \\
 &\Leftrightarrow k < (n+1)p
 \end{aligned}$$

There are two cases: if  $(n+1)p$  is not an integer, then  $k_0 = \lfloor (n+1)p \rfloor$  is the single value that has maximal weight, and if  $(n+1)p$  is an integer, then both  $k_0 = (n+1)p$  and  $k_0 - 1$  have maximal weight. With further analysis, one can show for any  $\epsilon > 0$  fixed,  $p$  fixed, as  $n \rightarrow \infty$ ,

## 2.2.2 Bernoulli's Law of Large Numbers

$$\mathbb{P}\left(\left|\frac{X}{n} - p\right| > \epsilon\right) = 0$$

This is called Bernoulli's Law of Large Numbers. This is effective for  $X \sim \text{Binom}(n, p)$ : fix  $p$ , and let  $n \rightarrow \infty$ .

- Ex. 2.2.3** (a) Shoot at a target 10 times, and suppose we hit the target with  $p = 0.1$ . Let  $X^{(a)}$  denote the number of hits, so that  $\mathbb{P}(X^{(a)} > 1) \approx 0.264\dots$
- (b) Shoot at a target 20 times, and suppose we hit the target with  $p = 0.05$ .
- (c) Shoot at a target 100 times, and suppose we hit the target with  $p = 0.01$ . Let  $X^{(c)}$  denote the number of hits, so that  $\mathbb{P}(X^{(b)} > 1) = 0.26424\dots$

## 2.3 Poisson Distribution

### 2.3.1 Basic Properties

**Def'n. 2.3.1** Let  $\lambda > 0$  be a parameter. Then  $X \sim \text{Poi}(\lambda)$  if it can take  $0, 1, 2, 3, \dots$  and  $\mathbb{P}(X = k)e^{-\lambda} \frac{\lambda^k}{k!}$ .

**Prop. 2.3.2** Let  $n \rightarrow \infty$ ,  $p = p(n) \rightarrow 0$  so that  $np(n) \rightarrow \lambda$ . Then for any  $k \in \mathbb{N}$ ,

$$\binom{n}{k} p^k (1-p)^{n-k} \rightarrow e^{-\lambda} \frac{\lambda^k}{k!}$$

**PROOF** Recall that

$$\lim \left(1 + \frac{1}{n}\right)^n = e, \quad \lim \left(1 - \frac{1}{n}\right)^n = e^{-1}$$

Thus let  $a(n) \rightarrow \infty$  such that  $\lim n/a(n) = x$ . Then

$$\lim \left(1 - \frac{1}{a(n)}\right)^n = \lim \left[ \left(1 - \frac{1}{a(n)}\right)^{a(n)} \right]^{\frac{n}{a(n)}} = e^{-x}$$

Now note that  $\lim(n-u)p(n) = \lambda$ ,  $\lim(1-p(n))^{-k} = 1$ , and

$$\lim(1-p)^n = \lim(1-p(n))^n = \lim \left(1 - \frac{1}{1/p(n)}\right)^n = e^{-\lambda}$$

so

$$\lim \frac{1}{k!} n(n-1)\cdots(n-k+1)p \cdot p \cdots p(1-p)^{-k}(1-p)^n = e^{-\lambda} \frac{\lambda^k}{k!} \quad \square$$

In words, many independent trials with small success rate can be approximated by the Poisson distribution. None

**Ex. 2.3.3** Since “poisson” means fish in french, we have an example about fishing. A fisherman goes fishing every day. He says: “on average, on 1/5 days, I do not catch any fish”. What is the chance that he catches at least two fish next time?

Let  $X$  count the number of fish on a particular occasion. There are many fish which move independently in the lake (maybe this assumption is not accurate, but we'll ignore that), and for any fixed fish, the chance of being caught is small. Thus we can assume  $X \sim \text{Poi}(\lambda)$ . Then  $X^{(1)}, X^{(2)}, \dots$  is the number of fish caught in consecutive occasions. Then we have

$$\frac{|\{X^{(k)} = 0 | k = 1, \dots, n\}|}{n} \approx \frac{1}{5}$$

so by the Bernoulli Law of Large Numbers,

$$\mathbb{P}(X = 0) = \frac{1}{5}$$

so  $e^{-\lambda} = 1/5$ . Then  $\lambda = \ln 5$  so that

$$\mathbb{P}(X \geq 2) = 1 - \mathbb{P}(X = 0) - \mathbb{P}(X = 1) = 1 - e^{-\lambda} - \lambda e^{-\lambda} \approx 0.48$$

If  $X \sim \text{Poi}(\lambda)$ , then

$$\mathbb{E}(X) = \sum_{k=0}^{\infty} k \cdot e^{-\lambda} \frac{\lambda^k}{k!} = \lambda e^{-\lambda} \sum_{m=0}^{\infty} \frac{\lambda^m}{m!} = \lambda e^{-\lambda} e^{\lambda} = \lambda$$

Similarly,

$$\mathbb{E}(X(X-1)) = \sum_{k=0}^{\infty} k(k-1) e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} \lambda^2 \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} = \lambda^2$$

so that  $\mathbb{E}(X^2) = \lambda^2 + \lambda$  and  $\text{Var}(X) = \lambda$ .

**Ex. 2.3.4** How many chocolate chips should you add per muffin so that no more than 1% of muffins have no chocolate chips in them?

Let  $X$  denote the number of chips in a given muffin is poisson distributed, so  $X \sim \text{Poi}(\lambda)$ . As well,  $\lambda = N \cdot 1/M$ , so  $1/M$  is the success rate (the probability that a given chip is in the given muffin). We need  $\mathbb{P}(X = 0) \leq 0.01$ , so  $e^{-\lambda} \leq 0.01$  and  $\lambda \geq \ln 100 \approx 4.6$ .

## 2.3.2 Poisson Process

A stochastic process is a random phenomenon evolving in time. For example, a point process is a random collection of points on  $[0, +\infty)$ . There are three properties that characterize the Poisson process.

**Def'n. 2.3.5** Consider two integer valued random variables, say  $X$  and  $Y$ . These are **independent** if for any  $k, l \in \mathbb{Z}$ , the events  $\{X = k\}$  and  $\{Y = l\}$  are independent.

**Def'n. 2.3.6** Let  $f, g$  be arbitrary functions with  $g(h) \rightarrow 0$  as  $h \rightarrow 0$ . Then  $f : \mathbb{R}^+ \rightarrow \mathbb{R}$  is said to be  $o(g)$  if  $\lim_{h \rightarrow 0^+} \frac{f(h)}{g(h)} = 0$

We thus have the following properties characterizing the Poisson process with intensity  $\lambda$ .

1. Let  $I_1, I_2, \dots, I_n$  be non-overlapping intervals in  $[0, \infty)$ . Let  $X_k$  denote the number of points in  $I_k$  for  $k = 1, \dots, n$ . Then the events  $X_1, X_2, \dots, X_n$  are independent.
2. Homogeneity. Let  $I$  be an interval of length  $h$ . Then  $\mathbb{P}(\text{there is exactly one point in } I) = \lambda h + o(h)$ .
3. No accumulations. Let  $I$  be an interval of length  $h$ . Then  $\mathbb{P}(\text{there are at least two points in } I) = o(h)$ .



**Prop. 2.3.7** Let a point process satisfy the above properties. Then if  $I_t$  is an arbitrary interval of length  $t$  and  $N_t$  denotes the number of impacts with  $I_t$ , then  $N_t \sim \text{Poi}(\lambda t)$ .

**PROOF** We want to compute  $\mathbb{P}(N_t = k)$ . We will consider this using an additional parameter  $n$ . Without loss of generality, consider the interval  $[0, t]$ . Define

$$I_i^{(n)} = \left[ \frac{(i-1)t}{n}, \frac{it}{n} \right), \quad i = 1, \dots, n$$

Let  $E_i^{(n)}$  denote the event in which there is exactly one point in  $I_i^{(n)}$ , and  $D_i^{(n)}$  is the event in which there are at least two points in  $I_i^{(n)}$ . Let  $D = D^{(n)} = \bigcup_{i=1}^n D_i^{(n)}$ . Then

$$\mathbb{P}(D) \leq \sum_{i=1}^n \mathbb{P}(D_i^{(n)}) = n \cdot o(t/n) = \frac{o(t/n)}{t/n} \cdot t = o(1)$$

which tends to 0 as  $n$  goes to infinity. But then

$$\begin{aligned} \mathbb{P}(N_t = k) &= \mathbb{P}(\{N_t = k\} \cap D) + \mathbb{P}(\{N_t = k\} \cap D^c) = o(1) + \binom{n}{k} p(n)^k (1-p)^{n-k} \\ &= e^{-\lambda t} \frac{(\lambda t)^k}{k!} \end{aligned}$$

as  $n$  goes to infinity. Note that  $\{N_t = k\} \cap D^c$  means that there are exactly  $k$  out of the  $E_1^{(n)}, \dots, E_n^{(n)}$  that occur, i.e. there are  $k$  out of  $n$  independent trials. As well,  $p(n) = \mathbb{P}(E_1^{(n)}) = \lambda t/n + o(t/n)$ , so  $n \cdot p(n) = \lambda t + o(1)$ .  $\square$

**Ex. 2.3.8** On average, there are two earthquakes per week. What is the probability that there are at least 3 earthquakes in the first two weeks? How much time elapses until the first earthquake occurs?

The intensity of the process is given by  $\lambda = 2$  per week, so in 2 weeks,  $N_2 \sim \text{Poi}(4)$ . Then  $\mathbb{P}(N_2 \geq 3) = 1 - \mathbb{P}(N_2 \leq 2) = 1 - e^{-4} - 4e^{-4} - \frac{4^2}{2}e^{-4}$ .

Now let  $T$  denote the time that elapses until the first earthquake. Fix some  $t > 0$ . This amounts to computing  $\mathbb{P}(T > t)$ .  $T$  is a continuous random variable, and note that

$$\mathbb{P}(T > t) = \mathbb{P}(N_t = 0) = e^{-\lambda t}$$

Here are two more constructions.

**Ex. 2.3.9** 1. Consider a Poisson process of intensity  $\lambda$ . Each point is colored white with probability  $p$  and orange with probability  $1 - p$ . The collection of orange points make a Poisson process with intensity  $(1 - p)\lambda$ . This can be proven by considering a Poisson distributed random variable, and for each  $k$ , think of a collection of  $k$  objects that you keep with probability  $p$ . Then the number of points you get is still Poisson distributed.

2. If  $X_1 \sim \text{Poi}(\mu_1)$  and  $X_2 \sim \text{Poi}(\mu_2)$ , then  $X_1 + X_2 \sim \text{Poi}(\mu_1 + \mu_2)$ .

**Ex. 2.3.10** Let  $X$  denote the number of matches. We have  $\mathbb{P}(X = 0) = 1 - 1 + \frac{1}{2} - \frac{1}{3!} + \cdots \pm \frac{1}{N!}$  which tends to  $1/e$  for large  $k$ . Now, for some  $j < N$ , we claim that

$$\mathbb{P}(X = j) = \binom{N}{j} \mathbb{P}(\text{No matches for } N - j \text{ people}) \cdot \frac{1}{N} \frac{1}{N-1} \cdots \frac{1}{N-j+1}$$

and fix  $j$ , and consider the limit as  $N$  goes to infinity, yielding  $\frac{1}{j!}e^{-1}$ . Thus

$$\mathbb{P}(Y = j) = \frac{e^{-1}}{j!}$$

To see why this works, let  $A_i$  denote the event in which person  $i$  is matched with her phone. We have  $\mathbb{P}(A_j) = \frac{1}{N}$ . However,  $\mathbb{P}(A_1|A_2) = \frac{1}{N-1}$ , so as  $N \rightarrow \infty$ , the events are almost pairwise independent.

## 2.4 Additional Discrete Distributions

### 2.4.1 Geometric Distribution

**Def'n. 2.4.1** A *geometric distribution* is a sequence of independent trials, with  $p \in (0, 1)$  success rate. Then  $\mathbb{P}(X = k) = p(1 - p)^{k-1}$ .

We certainly have  $\sum_{k=1}^{\infty} q^{k-1}p = 1$ . Now let

$$g(q) = \frac{1}{1-p} = \sum_{k=0}^{\infty} q^k$$

If  $X \sim (p)$ , then

$$\begin{aligned} \mathbb{E}(X) &= \sum_{k=0}^{\infty} kq^{k-1}p \\ &= p \sum_{k=1}^{\infty} \frac{d}{dq} q^k \\ &= p \frac{d}{dq} \left( \sum_{k=0}^{\infty} q^k \right) \\ &= pg'(q) = \frac{1}{p} \end{aligned}$$

and by similar methods,

$$\mathbb{E}(X(X-1)) = pqg''(q) = 2 \frac{(1-p)}{p^2}$$

so that

$$\mathbb{E}(X^2) = \frac{2q+p}{p^2} = \frac{1+q}{p^2}$$

and

$$\text{Var}(X) = \frac{q}{p^2}$$

The geometric distribution also has the memoryless property. Note  $\mathbb{P}(X > k) = q^k$  and  $\mathbb{P}(X > k + n) = q^{k+n}$  so that  $\mathbb{P}(X > k + n | X > n) = q^k$ .

## 2.4.2 Negative Binomial Distribution

**Def'n. 2.4.2** A *negative binomial distribution* has parameters  $p \in (0, 1)$ ,  $r \geq 1$ , and  $X$  can take  $r, r + 1, \dots$  where  $\{X = k\}$  is the event in which success  $r$  occurs at trial  $k$ .

$\mathbb{P}(X = r) = p^r$ ,  $\mathbb{P}(X = r + 1) = rp^r q$ . In general,

$$\mathbb{P}(X = k) = \binom{k-1}{r-1} p^r q^{k-r}$$

**Ex. 2.4.3** This is Stefan Banach's matchbox problem. There are  $N$  matches in two boxes. At every stage, he randomly takes a match from either box. What is the probability that when he runs out of matches in one of the matchboxes, there are exactly  $k$  matches in the other box?

Will do this later.

We have

$$\mathbb{E}(X) = \frac{r}{p}, \quad \text{Var } X = r \frac{q}{p^2}$$

## 2.4.3 Hypergeometric distribution

Consider parameters  $N, M < N, n < N$  and there are  $N$  balls,  $M$  black,  $N - M$  white. We draw  $n$  balls without replacement. We want  $\{X = k\}$  to be the event that there are exactly  $k$  black balls in  $n$  balls drawn. Then

$$\mathbb{P}(X = k) = \frac{\binom{M}{k} \cdot \binom{N-M}{n-k}}{\binom{N}{n}}$$

One can compute

$$\mathbb{E}(X)M \cdot \frac{n}{M}$$

Let  $X$  and  $Y$  be arbitrary discrete random variables. Then  $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$ .

**PROOF** We have

$$\begin{aligned} \mathbb{E}(X + Y) &= \sum_{k=1}^{\infty} \sum_{l=1}^{\infty} (x_k + y_l) \mathbb{P}(X = x_k, Y = y_l) \\ &= \sum_{k=1}^{\infty} x_k \sum_{l=1}^{\infty} \mathbb{P}(X = x_k, Y = y_l) + \sum_{l=1}^{\infty} y_l \sum_{k=1}^{\infty} \mathbb{P}(X = x_k, Y = y_l) \\ &= \sum_{k=1}^{\infty} x_k \mathbb{P}(X = x_k) + \sum_{l=1}^{\infty} y_l \mathbb{P}(Y = y_l) \\ &= \mathbb{E}(X) + \mathbb{E}(Y) \end{aligned}$$

□

Now suppose  $X$  and  $Y$  are independent. Then

$$\begin{aligned}\mathbb{E}(XY) &= \sum_{k=1}^{\infty} \sum_{l=1}^{\infty} x_k y_l \mathbb{P}(X = x_k, Y = y_l) \\ &= \sum_{k=1}^{\infty} \sum_{l=1}^{\infty} x_k y_l \mathbb{P}(X = x_k) \mathbb{P}(Y = y_l) \\ &= \left( \sum_{k=1}^{\infty} x_k \mathbb{P}(X = x_k) \right) \left( \sum_{l=1}^{\infty} y_l \mathbb{P}(Y = y_l) \right) \\ &= \mathbb{E}(X) \mathbb{E}(Y)\end{aligned}$$

so, if  $X$  and  $Y$  are independent, then  $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$ . We can use this to compute the expected values in previous examples.

1. Suppose  $X \sim \text{Binom}(n, p)$ . Then  $X = \eta_1 + \eta_2 + \dots + \eta_n$  where  $\eta_j = 1$  if the trial  $j$  is successful, and 0 otherwise. As well,  $\mathbb{E}(\eta_j) = p$  and  $\text{Var}(\eta_j) = p(1 - p)$  so that  $\mathbb{E}(X) = np$  and  $\text{Var}(X) = np(1 - p)$ .
2. Suppose  $X \sim (N, M, n)$ . Then  $X = \eta_1 + \eta_2 + \dots + \eta_M$  where  $\eta_j = 1$  if ball  $j$  is drawn, and 0 otherwise. Then  $\mathbb{E}(\eta_j) = \mathbb{E}(\eta_1) = \mathbb{P}(\eta_1 = 1) = \frac{n}{M}$  and the result follows. However, these events are not independent, so we can't do the same thing for the variance.

Consider as well the negative binomial  $\{X = k\}$  is the event that event  $r$  occurs at trial  $k$ . Let  $T_i$  denote the number of trials after the  $(j - 1)^{\text{st}}$  success needed to obtain the  $j^{\text{th}}$  success. Then  $T_j$  are all independent and  $T_i \sim \text{Geom}(p)$  and

$$\mathbb{E}(T_j) = \frac{1}{p}, \quad \text{Var}(T_j) = \frac{q}{p^2}$$

so that

$$\mathbb{E}(X) = \frac{r}{p}, \quad \text{Var}(X) = \frac{rq}{p^2}$$

**Ex. 2.4.4** Consider  $M$  types, each item can be any of the types with  $1/M$  chance independently. Then if  $X$  denotes the number of items to collect all the type, then  $X = \sum Y_j$ , where  $Y_j$  is the number of items needed to collect type  $j$ . Since  $Y_1, \dots, Y_m$  are independent and  $Y_j \sim ((M - j + 1)/M)$ , we have

$$\mathbb{E}(X) = \sum_{j=1}^m \mathbb{E}(Y_j) = M \left( \frac{1}{M} + \frac{1}{M-1} + \dots + 1 \right)$$

**Ex. 2.4.5**  $M = 5$  people board the elevator on a  $K + 1 = 11$  storied building (on the first floor). Each of them picks a destination, one of the floors 2 through  $K + 1$ . Let  $X$  denote the number of times the elevator stops.

Define  $\eta_j$  to be 1, if the elevator stops on floor  $(j + 1)$ , and 0 otherwise. Then

$$\mathbb{E}(\eta_j) = \mathbb{E}(\eta_1) = 1 - \left( \frac{K-1}{K} \right)^M$$

so that

$$\mathbb{E}(X) = K \left( 1 - \left( \frac{K-1}{K} \right)^M \right)$$



# Chapter 3

## Continuous Random Variables

### 3.1 Cumulative Distribution Function

#### 3.1.1 Properties of the CDF

**Def'n. 3.1.1** Let  $X$  be an arbitrary random variable. Then the **cumulative distribution function** of  $X$  is  $F_X : \mathbb{R} \rightarrow \mathbb{R}$ , defined by

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

**Ex. 3.1.2** Let  $X$  denote the number of heads when 3 fair coins are flipped. Then  $\mathbb{P}(X = 0) = \mathbb{P}(X = 3) = 1/8$  and  $\mathbb{P}(X = 1) = \mathbb{P}(X = 2) = 3/8$ .

$$F_X(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{8} & 0 \leq x < 1 \\ \frac{1}{2} & 1 \leq x < 2 \\ \frac{7}{8} & 2 \leq x < 3 \\ 1 & 3 \leq x \end{cases}$$

In general, if  $X$  is a discrete random variable with no limit points, then  $F_X(x)$  is a step function.

**Ex. 3.1.3** Consider a Poisson process with intensity  $\lambda$ , and let  $T$  denote the time that elapses until the first impact occurs. Then

$$F_T = \begin{cases} \mathbb{P}(T \leq x) = 0 \\ 1 - \mathbb{P}(T > x) = 1 - \mathbb{P}(\{N_x = 0\}) = 1 - e^{-\lambda x} \end{cases} \quad x \leq 0$$

First note if  $a \leq b$ , then  $\mathbb{P}(A < x \leq b) = F_X(b) - F_X(a)$ .

**Prop. 3.1.4** For a cumulative distribution function, the following hold:

1.  $F(x)$  is non-decreasing
2.  $\lim_{x \rightarrow +\infty} F(x) = 1$
3.  $\lim_{x \rightarrow -\infty} F(x) = 0$

4.  $F(x)$  is right-continuous. Write this notation as  $F(x_0 + 0) = F(x_0)$ .

Furthermore, these properties characterize the cumulative distribution function: any function with these 4 properties is the distribution function of some random variable.

PROOF 1. Obvious.

2. By (1), it suffices to show that  $\lim_{x \rightarrow +\infty} F(x) = \lim_{n \rightarrow \infty} F(n)$ . Let  $E_n = \{X \leq n\}$ . Then  $E_n \subset E_{n+1}$  and  $\bigcup_{n=1}^{\infty} E_n = \Omega$ , so

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

3. Let  $F_n = \{X \leq n-\}$ , so  $F_n \supset F_{n+1}$  and  $\bigcap_{n=1}^{\infty} F_n = \emptyset$ . Then

$$\lim_{n \rightarrow -\infty} F(n) = \lim_{n \rightarrow -\infty} F(-n) = \lim_{n \rightarrow \infty} \mathbb{P}(E_n) = \mathbb{P}\left(\bigcup_{n=1}^{\infty} E_n\right) = \mathbb{P}(\Omega) = 1$$

4. By (1), we have

$$\begin{aligned} F(x_0 + 0) &= \lim_{n \rightarrow \infty} \mathbb{P}\left(\left\{X \leq x_0 + \frac{1}{n}\right\}\right) \\ &= \lim_{n \rightarrow \infty} \mathbb{P}(B_n) \\ &= \lim_{n \rightarrow \infty} \mathbb{P}(B_n) = \mathbb{P}\left(\bigcap_{n=1}^{\infty} B_n\right) \\ &= \mathbb{P}(X \leq x_0) \\ &= F(x_0) \end{aligned}$$

Note that  $F(x_0 - 0)$  may be less than  $F(x_0)$ . In fact,  $F(x_0) - F(x_0 - 0) = \mathbb{P}(X = x_0)$ .  $\square$

**Def'n. 3.1.5** A random variable  $X$  is **absolutely continuous** if there exists some  $f_X : \mathbb{R} \rightarrow \mathbb{R}$  such that for any Borel set  $V \subset \mathbb{R}$ ,

$$\mathbb{P}(X \in V) = \int_V f(x) dx$$

In particular,  $0 = \int_{x_0}^{x_0} f(x) dx = \mathbb{P}(\{x = x_0\})$ . As well,  $f(x) \geq 0$  for all  $x \in \mathbb{R}$  and  $\int_{-\infty}^{\infty} f(x) dx = 1$ . Moreover,

$$F_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^x f(t) dt$$

so  $f_X$  is an anti-derivative of  $F_X$ . Furthermore,

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

Note that it is not only true that  $F_X(x)$  is continuous; it is also differentiable almost everywhere, and  $\frac{dF}{dx} = f(x)$  almost everywhere.

**Ex. 3.1.6** Recall

$$F_T(t) = \begin{cases} 0 & : t \leq 0 \\ 1 - e^{-\lambda t} & : t > 0 \end{cases}$$

and

$$f_T(t) = \begin{cases} 0 & : t < 0 \\ \lambda e^{-\lambda t} & : t > 0 \end{cases}$$

If  $X$  is absolutely continuous, then  $\mathbb{P}(X = a) = 0$ . Then the values of  $f(x)$  not probabilities: consider  $a \in \mathbb{R}$ ,  $\epsilon \rightarrow 0$ . Then

$$\mathbb{P}(a \leq X \leq a + \epsilon) = \int_a^{a+\epsilon} f(x) dx \approx f(a)\epsilon + o(\epsilon)$$

is the best linear approximation of the cumulative distribution at  $a$ .