Course Notes

Introduction to Probability

Alex Rutar

BSM Fall 2018

Contents

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 Ω

Ω 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 $\Omega = \{(a,b,c),(a,c,b),\ldots,(c,b,a)\}$. 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 $\Omega = \{(1,1),(1,2),\ldots,(6,6)\}$ and $|\Omega| = 36$.

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

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

Then define

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

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 Ω, which denote the events that we consider.

- If Ω is countable, then typically $\mathcal{F}$ is just the collection of all subsets of Ω.
- If Ω 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 $A_1, A_2, \ldots \in \mathcal{F}$, then $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

in other words, that $\mathcal{F}$ is a σ−algebra.

1.1.4 $\mathbb{P}$

Finally, $\mathbb{P} : \mathcal{F} \to \mathbb{R}$ 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. (σ−additivity) Let $A_1, A_2, A_3, \ldots$ 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

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

  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 generlizes arbitrarily:

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

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

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

We can distribute the first and third terms using the induction hypothesis, and the result follows. □

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

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

Prop. 1.1.7 *σ−additivity implies that for any increasing sequence,*

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

and similarly for any decreasing sequence

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

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

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

and taking probabilities,

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

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

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

Thus

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

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

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

□

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 change 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 Ω 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

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

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 amoung 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 $B = A_1^c \cap A_2^c \dots \cap A_N^c = (A_1 \cup \dots \cup A_N)^c$, and $\mathbb{P}(A_i) = \frac{1}{N}$. Now in general,

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

for i_k distinct. Thus we now have

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

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

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

$$A_k = \{\text{the 8th woman sits next to her spouse}\}$$

so that

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

Note that

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

by "joining" the couple together, arranging them around the table, and permuting the couple internally. Thus generalizes to

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

Then by inclusion-exclusion,

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

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 $A, 2, 3, 4, \ldots, 10$.

Define $\Omega = \{5 \text{ element subsets of the 52 cards}\}$. 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 $\Omega = \{13 \text{ cards that North gets}\}$.

$$\mathbb{P}(\text{North receives all spades}) = \frac{1}{\binom{52}{13}}$$

$$\mathbb{P}(\text{North does not receive all 4 suits of any value}) = \\ 1 - \mathbb{P}(\text{There is some value such that all suits are at N})$$

Let $V_k = \{\text{North gets all four suits of value } k\}$. Then

$$\mathbb{P}(V_1) = \frac{\binom{48}{9}}{\binom{52}{13}}$$

$$\mathbb{P}(V_1 \cap V_2) = \frac{\binom{44}{5}}{\binom{52}{13}}$$

$$\mathbb{P}(V_1 \cap V_2 \cap V_4) = \frac{\binom{40}{1}}{\binom{52}{13}}$$

Thus

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

What is the change 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 even 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 $\mathbb{P}(\cdot|E) : \mathcal{F} \to \mathbb{R}$. 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 $F_1, F_2, \ldots$ are mutually exclusive, then

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

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

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

Proof This follows by induction from the definition of conditional probability. □

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.

1. What is $\mathbb{P}$(the first three balls are red, blue, red (in this order)).

Sol'n 1. Let R_i, B_i denote the i^{th} draw is red or blue respectively. Then

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

Ex. 1.2.6 What is $\mathbb{P}$(in bridge, each of the players gets one ace)?

Sol'n Write

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

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 compacy, 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-addicent-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

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

Prop. 1.2.8 *Suppose $A_1, A_2, \ldots, A_n \in \mathcal{F}$ form a partition of Ω. 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 ito 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

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

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 $A_i = \{\text{door } i \text{ hides the price}\}$, 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

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

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 $A_1, A_2, \ldots$ are* ***independent (as a collection)*** *if, for any choice of indices $1 \leq i_1 < i_2 < \cdots < i_k \leq n$, then*

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

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 $A_k = \{\text{the } k^{\text{th}} \text{ trial is a succes}\}$ so that $\mathbb{P}(A_k) = p$ and the events $A_1, \ldots, A_n$ are independent. Our framework is to consider the space $\Omega \times \Omega \times \cdots \times \Omega$.

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

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

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

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

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,

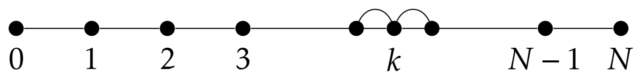
$$\begin{aligned}\mathbb{P}(D) &= \mathbb{P}(D|A_1)\mathbb{P}(A_1) + \mathbb{P}(D|A_2)\mathbb{P}(A_2) + \mathbb{P}(D|C_1)\mathbb{P}(C_1)\\ &= \mathbb{P}(B_1) + \mathbb{P}(C_1)\mathbb{P}(D)\end{aligned}$$

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

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

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

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

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 $i = 0,\ldots,k$). 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 $H_j = \{\text{the } j^{\text{th}} \text{ flip is H}\}$ for $j = 1, 2, \ldots, n, n+1$. Then the events H_j are not independent, but they are conditionally independent given any of the $C_i = \{\text{the } i^{\text{th}} \text{ coin is initially picke}\}$

for $i = 0,\ldots,k$. Moreover, $\mathbb{P}(H_j|C_k) = \frac{i}{k}$. We thus have

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

Both the numerator and denominator of $p(k,n)$ are sums of the form $\sum\limits_{i=0}^{k} f(i/k)\cdot 1/k$. Thus as $k \to \infty$,

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

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 Define a k–strategy for each $k = 1,\ldots,N$, in which we observe the first k items, and pick the first of the remaining ones that is better than the first k. Define

$$P_k^{(N)} = \mathbb{P}(\text{choose the best with the } k\text{–strategy})$$

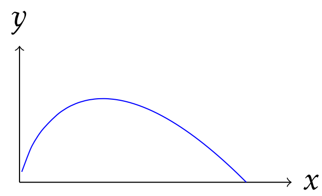
Let B_k denote this event and A_i be the event that the best prize is at the i^{th} position, so $\mathbb{P}(A_i) = 1/N$. Note that $\mathbb{P}(B_k|A_j) = 0$ for $j \le k$, and $\mathbb{P}(B_k|A_j) = \frac{k}{j+1}$ for $j > k$. Then

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

We can then compute

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

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

Random Variables

2.1 Basics

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

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 $Y : \Omega \to \{0, 1, 2, 3\}$.
2. Repeatedly roll a fair die until a 6 occurs. Let Z denote the number of rolls necessary. Now $Z : \Omega \to \mathbb{N}$.

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 $p : \mathbb{R} \to \mathbb{R}$ 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 funtion 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 $\{X > n\} = A_1^{(n)} \cup \cdots \cup A_k^{(n)}$ where $A_k^{(n)}$ is the event that type k has not been collected in n items.

Now

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

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

$$\mathbb{P}(X = 0) = \frac{1}{2}, \quad \mathbb{P}(X = 100) = \frac{1}{2} \cdot \frac{5}{6}, \quad \mathbb{P}(X = 100 + x) = \frac{1}{2} \cdot 16$$

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

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

so

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

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 $x_1, x_2, \ldots$ denote the possible values of X, and $y_1, y_2, \ldots$ denote the possible values of Y. Then

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

□

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

Proof Follows from linearity of the sum. □

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 $Pr(X = -100) = 1/2$. 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* $(X) = \mathbb{E}[(X - \mu)^2]$.

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

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

Prop. 2.1.11 $(aX + b) = a^2 X$.