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 $\bigcup\limits_{i=1}^{\infty} A_i \in \mathcal{F}$.

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 $D_1, D_2, \dots$ 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 $\lim_{n\to\infty} D_n = \bigcap_{n=1}^{\infty} D_n$ 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 $I_1, I_2, \dots$ 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

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

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

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

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 $(i_1,\ldots,i_{20}) \sim (i_{20}, i_1, \ldots, i_{19})$. 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 $\Omega = \{\text{5 element subsets of the 52 cards}\}$. 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 $\Omega = \{\text{13 cards that North gets}\}$.

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

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

So that, by Inclusion Exclusion,

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

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

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

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. 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

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

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

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

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 $A_i = \{\text{door } i \text{ hides the price}\}$, 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 $S = \bigcap\limits_{n=1}^{\infty} S_n$ 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,

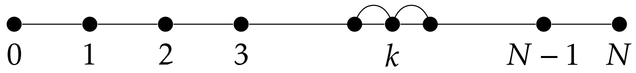
$$\begin{aligned}\mathbb{P}(D) &= \mathbb{P}(D|A_1)\mathbb{P}(A_1) + \mathbb{P}(D|A_2)\mathbb{P}(A_2) + \mathbb{P}(D|C_1)\mathbb{P}(C_1)\\ &= \mathbb{P}(B_1) + \mathbb{P}(C_1)\mathbb{P}(D)\end{aligned}$$

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 $A = \{$the patient is sick$\}$ has $\mathbb{P}(A) = 0.005$ so $\mathbb{P}(A^c) = 0.995$. Let $B_1 = \{$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 $B_2 = \{$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 $B_Y = \{$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 $i = 0, \dots, k$). 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^{th} flip is H for $j = 1, 2, \dots, n, n+1$. Let C_i denote the event in which the i^{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

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

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

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

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 We will solve this over a fixed type of strategy. 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. 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^{th} position, so $\mathbb{P}(A_i) = 1/N$.

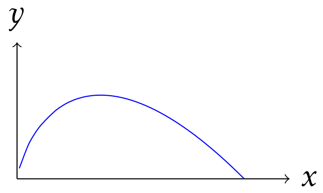
Now, $\mathbb{P}(B_k|A_j) = 0$ for $j \le k$, and $\mathbb{P}(B_k|A_j) = \frac{k}{j+1}$ for $j > k$. Then

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

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

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

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

Discrete 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 $\mathbb{E}(X) = \sum\limits_{k=1}^{\infty} x_k \mathbb{P}(X = x_k)$.*

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 $\mathbb{E}(X) = \sum\limits_{k=1}^{\infty} x_k \mathbb{P}(X = x_k)$ 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 $\mathbb{E}(g(X)) = \sum\limits_{k=1}^{\infty} g(x_k)\mathbb{P}(X = x_k)$

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

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

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

Proof Follows from linearity of the sum. □

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* $\mathrm{Var}(X) = \mathbb{E}[(X - \mu)^2]$.

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

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

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

Ex. 2.1.12

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

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

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 $\mathrm{Var}(\eta) = p - p^2 = pq$.

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

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

We can also compute

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

so that

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

and

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

We thus say that $X = \eta_1+\eta_2+\cdots+\eta_n$ 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 $X \sim \mathrm{Binom}(n,p)$, so that

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

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 $n \to \infty$,

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 $X \sim \mathrm{Binom}(n,p)$: fix p, and let $n \to \infty$.

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 $\mathbb{P}(X^{(a)} > 1) \approx 0.264\ldots$.

(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 $\mathbb{P}(X^{(b)} > 1) = 0.26424\ldots$

2.3 Poisson Distribution

2.3.1 Basic Properties

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

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

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

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 $a(n) \to \infty$ 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

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

In words, many independent trials with small success rate can be approximated by the Poisson distribution.

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 atches 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 $X \in \mathrm{Poi}(\lambda)$. Then $X^{(1)}, X^{(2)}, \ldots$ is the number of fish caught in conseqūtive occasions. Then we have

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

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 $X\sim\mathrm{Poi}(\lambda)$, 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,

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

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

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 $X\sim\mathrm{Poi}(\lambda)$. 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 $g(h)\to 0$ as $h\to 0$. Then $f:\mathbb{R}^+\to\mathbb{R}$ is said to be $o(g)$ of* $\lim_{h\to 0^+}\frac{f(h)}{g(h)}=0$

We thus have the following properties characterizing the Poisson process with intensity λ.

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. Homogenity. 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 $N_t \sim \operatorname{Poi}(\lambda t)$.*

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 occurr, 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 $n \cdot p(n) = \lambda t + o(1)$. □

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 occurrs?

The intensity of the process is given by $\lambda = 2$ per week, so in 2 weeks, $N_2 \sim \operatorname{Poi}(4)$. 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 λ. 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 $X_1 \sum \operatorname{Poi}(\mu_1)$ and $X_2 \sim \operatorname{Poi}(\mu_2)$, then $X_1 + X_2 \sim \operatorname{Poi}(\mu_1 + \mu_2)$.

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 $N \to \infty$, 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 $X \sim \text{Geom}(p)$, then

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

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

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

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 $r, r+1, \ldots$ 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

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

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 $\mathrm{Var}(X+Y) = \mathrm{Var}(X)+\mathrm{Var}(Y)$. We can use this to compute the expected values in previous examples.

1. Suppose $X \sim \mathrm{Binom}(n,p)$. Then $X = \eta_1 + \eta_2 + \cdots + \eta_n$ where $\eta_j = 1$ if the trial j is successful, and 0 otherwise. As well, $\mathbb{E}(\eta_j) = p$ and $\mathrm{Var}(\eta_j) = p(1-p)$ so that $\mathbb{E}(X) = np$ and $\mathrm{Var}(X) = np(1-p)$.

2. Suppose $X \sim \mathrm{Hypergeometric}(N,M,n)$. Then $X = \eta_1 + \eta_2 + \cdots + \eta_M$ where $\eta_j = 1$ if back ball j is drawn, and 0 otherwise. Then $\mathbb{E}(\eta_j) = \mathbb{E}(\eta_1) = \mathbb{P}(\eta_1 = 1) = \frac{n}{N}$ 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 $(j-1)^{\mathrm{st}}$ success needed to obtain the j^{th} success. Then T_j are all independent and $T_i \sum \mathrm{Geom}(p)$ and

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

so that

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

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 fo collect all the type, then $X = \sum Y_j$, where Y_j is the number of items needed to collect type j. Since $Y_1,\ldots,Y_m$ are independent and $Y_j \sim \mathrm{Geom}((M-j+1)/M)$, we have

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

Ex. 2.4.5 $M = 5$ people baord 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 η_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 $F_X : \mathbb{R} \to \mathbb{R}$, 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 λ, and let T denote the time that elapses until the first impact occurs. Then

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

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. $\lim_{x\to+\infty} F(x) = 1$

3. $\lim_{x\to-\infty} F(x) = 0$

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 $\lim_{x\to+\infty}F(x)=\lim_{n\to\infty}F(n)$. Let $E_n=\{X\leq n\}$. Then $E_n\subset E_{n+1}$ and $\bigcup\limits_{n=1}^{\infty}=\Omega$, so

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

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

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

4. By (1), we have

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

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

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

$$\mathbb{P}(X\in V)=\int_V f(x)\,\mathrm{d}x$$

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

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

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

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

Note that it is not only true that $F_X(x)$ is continuous; it is also differentiable almost everywhere, and $\frac{\mathrm{d}F}{\mathrm{d}x}=f(x)$ 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 $a \in \mathbb{R}$, $\epsilon \to 0$. Then

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

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

Def'n. 3.1.7 *The expected value of X is $\mathbb{E}(X) = \int_{-\infty}^{\infty} x f_X(x)\,\mathrm{d}x$, if the integral is absolutely convergent.*

Prop. 3.1.8 $\mathbb{E}(Y) = \mathbb{E}(g(X)) = \int_{-\infty}^{\infty} g(x) f_X(x)\,\mathrm{d}x$.

Proof Let Y be a non-negative random variable with density $f_Y(y)$. Then we will see that $\mathbb{E}(Y) = \int_0^\infty \mathbb{P}(Y > y)\,\mathrm{d}y$. Note that

$$\begin{aligned} \int_0^\infty \mathbb{P}(Y > y)\,\mathrm{d}y &= \int_0^\infty \int_y^\infty f_Y(x)\,\mathrm{d}x\,\mathrm{d}y \\ &= \iint_D H(x,y)\,\mathrm{d}x\,\mathrm{d}y \\ &= \int_0^\infty \int_0^x f_Y(x)\,\mathrm{d}y\,\mathrm{d}x \\ &= \int_0^\infty x f_Y(x)\,\mathrm{d}x \\ &= \mathbb{E}(Y) \end{aligned}$$

Without loss of generality, assume $g(x) \geq 0$ (or write $g(x) = g_+(x) - g_-(x)$). Thus $Y = g(X) \geq 0$, so

$$\begin{aligned} \mathbb{E}(g(X)) &= \mathbb{E}(Y) \\ &= \int_0^\infty \mathbb{P}(Y > y)\,\mathrm{d}y \\ &= \int_0^\infty \mathbb{P}(g(X) > y)\,\mathrm{d}y \\ &= \int_{-\infty}^\infty \int_0^{g(x)} f_X(x)\,\mathrm{d}y\,\mathrm{d}x \\ &= \int_{-\infty}^\infty g(x) f_X(x)\,\mathrm{d}x \end{aligned}$$

For any $y_0 > 0$, $\{(x,y) \in U | y = y_0\} = D_{y_0}$. Thus $(x,y) \in U$ if and only if $g(x) > y > 0$, so U is the domain between the x–axis and the graph of g. □

Consider the Cantor function. Suppose this function had a density function $f(x)$, we would have that $F'(x) = f(x)$ for almost every x. Conversely, $F'(x) = 0$ almost everywhere.

3.2 Important Absolutely Continuous Distributions

3.2.1 Uniform Distribution

We say $X \sim \mathrm{UNI}[\alpha,\beta]$ for $\alpha < \beta$ if

$$f_X(x) = \begin{cases} \frac{1}{\beta-\alpha} & : \alpha \leq x \leq \beta \\ 0 & : \text{otherwise} \end{cases}$$

so that

$$F_X(x) = \begin{cases} 0 & : x < \alpha \\ \frac{x-\alpha}{\beta-\alpha} & : \alpha \leq x \leq \beta \\ 1 & : x > \beta \end{cases}$$

Now if $Y \sim \mathrm{UNI}[0,1]$, then

$$\mathbb{E}(Y) = \int_0^1 x \,\mathrm{d}x = \frac{1}{2}, \quad \mathbb{E}(Y^2) = \int_0^1 x^2 \,\mathrm{d}x = \frac{1}{3}$$

so that

$$\mathrm{Var}(Y) = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$$

and in general,

$$\mathbb{E}(X) = (\beta-\alpha)\,\mathbb{E}(Y) + \alpha = \frac{\beta+\alpha}{2}$$
$$\mathrm{Var}(X) = (\beta-\alpha)^2\,\mathrm{Var}(Y) = \frac{(\beta-\alpha)^2}{12}$$

Ex. 3.2.1 Buses leave from every station every 15 minutes, starting at 7am. What is the chance that I have to wait at least 10 minutes if

1. My arrival time at the station is uniform between 7am and 8am?

3.2.2 Standard Normal Distribution

We say Z is standard normal if $Z \sim N(0,1)$ if it has density

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

Then its cumulative distribution function is given by

$$\Phi(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}x$$

cannot be expressed in terms of elementary functions. However, we do have $\Phi(0) = 1/2$, and $\Phi(-x) = 1 - \Phi(x)$. Let's show that $\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1$. We have

$$\begin{aligned} I^2 &= \int_{-\infty}^{\infty} e^{-x^2/2}\,\mathrm{d}x \cdot \int_{-\infty}^{\infty} e^{-y^2/2}\,\mathrm{d}y \\ &= \iint_{\mathbb{R}^2} e^{-(x^2+y^2)/2}\,\mathrm{d}x\,\mathrm{d}y \\ &= \int_0^{\infty}\int_0^{2\pi} e^{-r^2/2} r\,\mathrm{d}\theta\,\mathrm{d}r \\ &= 2\pi \int_0^{\infty} re^{-r^2/2}\,\mathrm{d}r = 2\pi \end{aligned}$$

as reqired. Furthermore, $\mathbb{E}(Z) = 0$ since $xf(x)$ is an odd function. As well,

$$\begin{aligned} \mathrm{Var}(Z) = \mathbb{E}(Z^2) &= \int_{-\infty}^{\infty} x^2 f(x)\,\mathrm{d}x \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 e^{-x^2/2}\,\mathrm{d}x \\ &= \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{2\pi}}\left(\left[-xe^{-x^2/2}\right]_0^{\infty} + \int_{-\infty}^{\infty} e^{-x^2}\,\mathrm{d}x\right) \\ &= \frac{1}{\sqrt{2\pi}} \cdot \sqrt{2\pi} = 1 \end{aligned}$$

Def'n. 3.2.2 *A random variable η is standard if $\mathbb{E}(\eta) = 0$ and $\mathrm{Var}(\eta) = 1$.*

Let X be an arbitrary random variable with $\mu = \mathbb{E}(X)$ and $\sigma^2 = \mathrm{Var}(X)$. Then $Y = (X - \mu)/\sigma$ is standard. In particular, Z is standard normal if $f_Z(x) = f(x) + \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ and $F_Z(x) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\,\mathrm{d}t$.

Def'n. 3.2.3 *Let $\mu \in \mathbb{R}$, $\sigma > 0$ be fixed parameters. Then $X \sim \mathcal{N}(\mu, \sigma^2)$ if $X = \sigma Z + \mu$, with $Z \sim \mathcal{N}(0,1)$.*

Ex. 3.2.4 The amount of beer in a class is normally distributed with expected value 0.5l with standard deviation 3cl. What is the chance that you get less than 4.5dl?

Sol'n Let's use dl everywhere. We have $X \sim \mathcal{N}(5, 0.09)$ since $\sigma = 0.3$. Then

$$\mathbb{P}(X \leq 4.5) = \mathbb{P}\left(\frac{X-\mu}{\sigma} \leq \frac{4.5-5}{0.3}\right) = \Phi(-1.66) = 1 - \Phi(1.66) \approx 0.0485$$

Thm. 3.2.5 (deMoivre-Laplace) *Let $X \sim \mathrm{Binom}(n,p)$ with p fixed and $n \to \infty$ and fix $a < b$. Then*

$$\mathbb{P}\left(a < \frac{X - np}{\sqrt{np(1-p)}} \leq b\right) \xrightarrow[n\to\infty]{} \Phi(b) - \Phi(a)$$

Proof (Stirling's Formula) For any $k \geq 1$,

$$0 \leq \int_k^{k+1} \ln(x)\,\mathrm{d}x - \frac{\ln k + \ln(k+1)}{2} \leq \frac{1}{k^2}$$

which follows by the Lagrange remainder. We then have

$$0 \leq \int_1^N \ln(x)\,\mathrm{d}x - \sum_{k=1}^{N-1} \frac{\ln k + \ln(k+1)}{2} \leq \sum_{k=1}^{N-1} \frac{1}{k^2}$$

where $\int_1^N \ln x\,\mathrm{d}x = N \ln N - N + 1$. As well,

$$\sum_{k=1}^{N-1} \frac{\ln k + \ln(k+1)}{2} = \ln(N!) - \ln(\sqrt{N})$$

so $\ln(N^N) - \ln(e^N) - \ln(N!) + \ln(\sqrt{N}) \to C$ as N goes to infinity. □

Ex. 3.2.6 Flip a coin 40 times. What is the probability that there are 20 heads?

We have $X \sim \mathrm{Binom}(40, 1/2)$, $\mathbb{E}(X) = 20$ and $\mathrm{Var}(X) = 10$. Then

$$\begin{aligned}
\mathbb{P}(X = 20) &= \mathbb{P}(19.5 < X \leq 20.5) \\
&= \mathbb{P}\left(\frac{19.5 - 20}{\sqrt{10}} < \frac{X - 20}{\sqrt{10}} \leq \frac{20.5 - 20}{\sqrt{10}}\right) \\
&= \mathbb{P}\left(\frac{-0.5}{\sqrt{10}} < Z \leq \frac{0.5}{\sqrt{10}}\right) \\
&= 2\Phi\left(\frac{0.5}{\sqrt{10}}\right) - 1 \\
&= 0.1272
\end{aligned}$$

and the true value is approximately 0.1254.

Ex. 3.2.7 There is a class with 400 students. Every student shows up at the lecture with 0.6 chance. How many seats are needed to ensure that everybody who shows up can sit down with 99% chance?

Let X denote the number of students who show up, so $X \sim \mathrm{Binom}(400, 0.6)$. Then $\mathbb{E}(X) = 240$ and $\sigma(X) = \sqrt{400 \cdot 0.6 \cdot 0.4} = \sqrt{96}$. We want

$$0.99 \leq \mathbb{P}(X \leq a) = \mathbb{P}\left(\frac{X - 240}{9.8} \leq \frac{9 - 240}{9.8}\right) = \Phi\left(\frac{a - 240}{9.8}\right)$$

and since $\Phi(2.33) \approx 0.9901$, we want $a - 240 \geq 2.33 \cdot 9.8$ and $a \geq\approx 2.63$.

Ex. 3.2.8 How many times does a fair coin have to be flipped to ensure that the proportion of heads is between 0.49 and 0.51 with 95% chance?

Let X denote the number of heads, with $X \sim \mathrm{Binom}(n, 1/2)$. We want

$$\begin{aligned} 0.95 &\leq \mathbb{P}(0.49 < X/n \leq 0.51) \\ &= \mathbb{P}(0.49n \leq X \leq 0.51n) \qquad\qquad = \mathbb{P}\left(\frac{0.49n - 0.51n}{\sqrt{n}/2} \leq \frac{X - n/2}{\sqrt{n}/2} \leq \frac{0.01n}{\sqrt{n}/2}\right) \\ &= \Phi(0.02\sqrt{n}) - (1 - \Phi(0.02\sqrt{n})) \end{aligned}$$

i.e. $\Phi(1.96) \approx 0.975 \leq \Phi(0.02\sqrt{n})$, in other words that $1.96 \leq 0.02\sqrt{n}$ and $n \geq 9604$.

$$0.96 \leq \mathbb{P}\left(\left|\frac{X}{n} - p\right| \leq 0.02\right) = \cdots$$

will get $Cp(1-p) \leq n$, and this holds for $C/4 \leq n$.

3.2.3 Exponential Distribution

Let $\lambda > 0$. Then $X \sim \mathrm{Exp}(\lambda)$ if its density is

$$f_X(x) = \begin{cases} 0 & : x < 0 \\ \lambda e^{-\lambda x} & : x \geq 0 \end{cases}$$

We thus have

$$\begin{aligned} \mathbb{E}(X^k) &= \int_0^\infty x^k \lambda e^{-\lambda x} \,\mathrm{d}x \\ &= \frac{k}{\lambda} \int_0^\infty x^{k-1} e^{-\lambda x} \,\mathrm{d}x \\ &\vdots \\ &= \frac{k!}{\lambda^k} \end{aligned}$$

In particular, $\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = 1/\lambda^2$ so $\sigma = 1/\lambda$. This has the memoryless property: if $X \sim \mathrm{Exp}(\lambda)$, and $t, s > 0$, then $\mathbb{P}(X > t) = 1 - \mathbb{P}(X \leq t) = e^{-\lambda t}$. Then

$$\mathbb{P}(X > t + s | X > s) = \frac{\mathbb{P}(X > s + t \cap X > s)}{\mathbb{P}(X > s)} = e^{-\lambda t}$$

Let X be a random variable with density $f_X(x)$ and distribution function $F_X(x)$. Let $Y = k(X)$ for some function $k : \mathbb{R} \to \mathbb{R}$.

Proof Suppose $k(x)$ is increasing. Then $F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(k(X) \leq y) = \mathbb{P}(X \leq k^{-1}(y)) = F_X(k^{-1}(y))$. By differentiation, $f_Y(y) = f_X(k^{-1}(y))\frac{\mathrm{d}}{\mathrm{d}y}(k^{-1}(y))$. The case when k is decreasing is analgous. □

Ex. 3.2.9 Suppose $H \sim \mathrm{UNI}[-\pi/2, \pi/2]$ and $k(x) = \tan(x)$. Then

$$f_X(x) = f_H(\tan^{-1}(x)) \cdot \left|\frac{\mathrm{d}}{\mathrm{d}x} \tan^{-1}(x)\right| = \frac{1}{\pi} \cdot \frac{1}{1 + x^2}$$

This distribution is called the standard Cauchy distribution and its expected value does not exist.

Def'n. 3.2.10 *The joint distribution (cumulative) function of X and Y is $F : \mathbb{R}^2 \to \mathbb{R}$ by*

$$F(x,y) = \mathbb{P}(X \le x, Y \le y)$$

Note that $\lim_{y\to\infty} F(x,y) = F_X(x)$, the marginal distribution function of X. Similarly, $\lim_{y\to-\infty} F(x,y) = 0$ for any fixed $x \in \mathbb{R}$. We also require $\mathbb{P}(a_1 < X \le b_1, a_2 < Y \le b_2) = F(b_1,b_2) - F(a_1,b_2) - F(b_1,a_2) + F(a_1,a_2) \ge 0$ as a probability on any rectangle.

Def'n. 3.2.11 *Let X, Y be discrete random variables. Then $p(x_k, y_l) = \mathbb{P}(X = x_k \cap Y = y_l)$ (and 0 otherwise).*

Ex. 3.2.12 Suppose there are 2 white, 2 red, and 1 blue ball in a box. Draw 2 balls, and let X count the number of red and Y count the number of blue. Then $p(x,y)$ is given by

X,Y	0	1	2	
0	$\frac{1}{10}$	$\frac{4}{10}$	$\frac{1}{10}$	$\frac{6}{10}$
1	$\frac{2}{10}$	$\frac{2}{10}$	0	$\frac{4}{10}$
	$\frac{3}{10}$	$\frac{6}{10}$	$\frac{1}{10}$	

Def'n. 3.2.13 *We say X and Y are jointly absolutely continuous if there exists $f : \mathbb{R}^2 \to R$ so that $\mathbb{P}((X,Y) \in D) = \iint_D f(x,y)\,\mathrm{d}x\,\mathrm{d}y$ for any Borel set $D \subset \mathbb{R}^2$.*

We thus have

$$F(a_1,a_2) = \int_{-\infty}^{a_1}\int_{-\infty}^{a_2} f(x,y)\,\mathrm{d}x\,\mathrm{d}y$$

so that

$$\frac{\partial F}{\partial x} = \int_{-\infty}^{a_2} f(a_1,y)\,\mathrm{d}y, \qquad \frac{\partial F^2}{\partial x \partial y}(x,y) = f(x,y) \ge 0$$

If (X,Y) are jointly absolutely continuous, then X is absolutely continuous and Y is absolutely continuous. Furthermore,

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

However, the converse can fail to hold:

Ex. 3.2.14 Say X is an arbitrary absolutely continuous variable, and $Y \equiv X$. But then $(X,Y) = (X,X)$ is restricted to a measure 0 subset, so no such density function $f(x,y)$ exists.

Ex. 3.2.15 Suppose X and Y have joint density

$$f(x,y) = \begin{cases} 2e^{-x}e^{-2y} & : x,y > 0 \\ 0 & \text{otherwise} \end{cases}$$

Then

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}y = \begin{cases} 0 & : x \le 0 \\ e^{-x} & : x > 0 \end{cases}$$

and

$$f_X(y) = \int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}y = \begin{cases} 0 & : x \le 0 \\ 2e^{-y} & : x > 0 \end{cases}$$

We can also compute

$$\begin{aligned}\mathbb{P}(X \geq Y) &= \int_{-\infty}^{\infty}\int_{-\infty}^{x} f(x,y)\,\mathrm{d}y\,\mathrm{d}x \\ &= \int_{0}^{\infty}\int_{0}^{x} 2e^{-2y}e^{-x}\,\mathrm{d}y\,\mathrm{d}x \\ &= \int_{0}^{\infty}(-e^{-3x}+e^{-x})\,\mathrm{d}x \\ &= \frac{2}{3}\end{aligned}$$

Def'n. 3.2.16 *Let D be a domain in $\mathbb{R}^2$ with finite measure. Then (X,Y) is uniformly distributed on D if their joint density satisfies*

$$f(x,y) = \begin{cases} 0 & : (x,y) \notin D \\ \frac{1}{\mu(D)} & : (x,y) \in D \end{cases}$$

Ex. 3.2.17 Let D denote the triangle with vertices $(0,0)$, $(3,0)$, and $(0,1)$. Then

$$f(x,y) = \begin{cases} \frac{2}{3} & : (x,y) \in D \\ 0 & : \text{otherwise} \end{cases}$$

so that

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}y = \begin{cases} 0 & : x \notin [0,3] \\ \frac{2}{3}\left(1-\frac{x}{3}\right) & : x \in [0,3] \end{cases}$$

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}y = \begin{cases} 0 & : y \notin [0,1] \\ 2-2y & : y \in [0,1] \end{cases}$$

Def'n. 3.2.18 *X and Y are independent if for any pair of intervals I,J,*

$$\mathbb{P}(X \in I, Y \in J) = \mathbb{P}(X \in I)\cdot\mathbb{P}(Y \in J)$$

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

In general, (X,Y) uniform on D are not independent unless D is a cartesian product of subsets of $\mathbb{R}$.

Ex. 3.2.19 Consider $f(x,y) = Axy$ whenever $x+y \leq$ and $x,y \in [0,1]$, and 0 otherwise. Then

$$f_X(x) = \begin{cases} 0 & : x \leq 0, x \geq 1 \\ \int_0^{1-x} Axy\,\mathrm{d}y = Ax(1-x)^2/2 \end{cases}$$

so that $A = 24$.

$$f_Y(y) = \begin{cases} 0 & : y \leq 0, y \geq 1 \\ \int_0^{1-y} Axy\,\mathrm{d}y = Ay(1-y)^2/2 \end{cases}$$

Ex. 3.2.20 A woman and a man agree to meet at some location between noon and 1pm. Both arrive at a random time, uniformly distributed within this hour independently. What is the probability that the one who arrives first has to wait more than 10 minutes?

Let $X \sim \text{UNI}[0,60]$, $Y \sim \text{UNI}[0,60]$. Then $\mathbb{P}(|X-Y|>10) = \frac{50^2}{60^2}$

Ex. 3.2.21 Consider a table ruled with parallel equidistant lines (2cm apart), on which we flip a needle of length 1cm. If X is the distance of the midpoint of the needle to the closest line, then $X \sim \text{UNI}[0,1]$. Let H denote the angle that the needle makes with the horizontal. Then

$$f(\theta,x) = \frac{2}{\pi}$$

whenever $0<x<1$ and $0<\theta<\pi/2$ so that

$$\begin{aligned}
\mathbb{P}(\text{intersection}) &= \mathbb{P}(X < \frac{1}{2}\cos\theta) \\
&= \iint_D f(x,\theta)\,\mathrm{d}x\,\mathrm{d}\theta \\
&= \frac{2}{\pi}\int_0^{\pi/2}\int_0^{\cos\theta/2} \mathrm{d}x\,\mathrm{d}\theta \\
&= \frac{1}{\pi}\int_0^{\pi/2} \cos\theta\,\mathrm{d}\theta \\
&= \frac{1}{\pi}
\end{aligned}$$

Let's now consider sums of independent random variables. If X and Y are independent with known distributions, what is the distribution of $X+Y$? If X,Y are discrete and integer valued,

$$\begin{aligned}
\mathbb{P}(X+Y=k) &= \sum_{l=-\infty}^{\infty} \mathbb{P}(X+Y=k|X=l)\mathbb{P}(X=l) \\
&= \mathbb{P}(Y=k-l)\mathbb{P}(X=l)
\end{aligned}$$

Suppose X has density $f_X(x)$ and Y has density $f_Y(y)$. What can we say about the distribution of $Z=X+Y$? We have $f(x,y)=f_X(x)f_Y(y)$ so that

$$\begin{aligned}
F_Z(z) &= \mathbb{P}(Z \le z) \\
&= \mathbb{P}(X+Y \le z) \\
&= \iint_H f(x,y)\,\mathrm{d}x\,\mathrm{d}y \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{z-x} f_X(x)f_Y(y)\,\mathrm{d}y\,\mathrm{d}x \\
&= \int_{-\infty}^{\infty} f_X(x)\left(\int_{-\infty}^{z-x} f_Y(y)\,\mathrm{d}y\right)\mathrm{d}x
\end{aligned}$$

and

$$f_Z(z) = \frac{\mathrm{d}F_z}{\mathrm{d}z}(z) = \int_{-\infty}^{\infty} f_X(x)f_Y(z-x)\,\mathrm{d}x$$

Suppose X and Y are independentt with densities $f_X(x)$ and $f_Y(y)$ respectively. Write $Z = X+Y$ and it has density

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

Ex. 3.2.22 Suppose $X, Y \sim \mathrm{UNI}[0,1]$. Then

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x)\,\mathrm{d}x = \begin{cases} 0 & : z \leq 0, z \geq 2 \\ z & : 0 < z \leq 1 \\ 2-z & 1 < z < 2 \end{cases}$$

Lemma 3.2.23 *Let $X \sim \mathcal{N}(0,1)$ and $Y \sim \mathcal{N}(0,\sigma^2)$ are independent, then $X + Y \sim \mathcal{N}(0, 1+\sigma^2)$.*

Proof First recall that $\mathbb{E}(X+Y) = \mathbb{E}(X) + \mathbb{E}(Y)$ and since X, Y are independent, $\mathbb{E}(X \cdot Y) = \mathbb{E}(X) \cdot \mathbb{E}(Y)$ so $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$. Thus it suffices to check that $X+Y$ is normally distributed. First note that

$$f_X(x) f_Y(z-x) = C_0 e^{-x^2/2} e^{-(z-x)^2/(2\sigma^2)} = C_0 e^{-(Ax^2 + Bzx + Dz^2)/2}$$

so we have

$$\begin{aligned} f_Z(z) &= C_0 e^{-Ez^2} \int_{-\infty}^{\infty} e^{-\frac{1}{2}A(x-Fz)^2}\,\mathrm{d}x \\ &= C_0 e^{-Ez^2} \int_{-\infty}^{\infty} e^{-\frac{A}{2}A(\tilde{x})^2}\,\mathrm{d}\tilde{x} \\ &= C_0 C_1 e^{-Ez^2} \end{aligned}$$

is normally distributed. □

Consider x_0, y_0 fixed and $\mathrm{d}y$ small. Then

$$\begin{aligned} \lim_{h \to 0} \mathbb{P}(X \leq x_0 | y_0 < y \leq y_0 + h) &= \lim_{h \to 0} \frac{\mathbb{P}(X \leq x_0 \cap Y \leq y_0 + h) - \mathbb{P}(X \leq y_0 \cap Y \leq y_0)}{\mathbb{P}(Y \leq y_0 + h) - \mathbb{P}(Y \leq y_0)} \\ &= \lim_{h \to 0} \frac{F(x_0, y_0 + h) - F(x_0, y_0)}{F_Y(y_0 + h) - F_Y(y_0)} \\ &= \lim_{h \to 0} \frac{\frac{F(x_0, y_0+h) - F(x_0, y_0)}{h}}{\frac{F_Y(y_0+h) - F_Y(y_0)}{h}} \\ &= \frac{\frac{\partial F}{\partial y}(x_0, y_0)}{F_Y'(y_0)} \\ &= \frac{\int_{-\infty}^{x_0} \frac{\partial^2 F}{\partial x \partial y}(t, y_0)\,\mathrm{d}t}{f_Y(y_0)} \\ &= \int_{-\infty}^{x_0} \frac{f(x,y)}{f_Y(y_0)}\,\mathrm{d}y \end{aligned}$$

This inspires the following definition:

Def'n. 3.2.24 *Let $y_0 \in \mathbb{R}$ be such that $f_Y(y_0) > 0$. Then the conditional density of X, given $Y = y_0$, is*

$$f_{X|Y}(X|Y = y_0) = \frac{f(x,y_0)}{f_Y(y_0)}$$

Ex. 3.2.25 Let $f(x,y)$ be uniform on the triangle with vertices $(0,0)$, $(0,1)$, and $(1,0)$. Determine the conditional distribution of Y given some $X = x_0$.

Firstly, we have

$$f_X(x_0) = \int_{-\infty}^{\infty} f(x_0,y)\,\mathrm{d}y = \begin{cases} \int_0^{x_0} 2\,\mathrm{d}y = 2x_0 & : 0 < x_0 < 1 \\ 0 & : \text{otherwise} \end{cases}$$

so that

$$f_{Y|X}(y|X = x_0) = \frac{f(x_0,y)}{f_X(x_0)} = \frac{f(x_0,y)}{2x_0} = \begin{cases} \frac{1}{x_0} & : 0 < y < x_0 \\ 0 & : \text{otherwise} \end{cases}$$

If (X,Y) are jointly uniformly distributed on some domain D, then the conditional distributions are always uniform on the relevant intervals. The marginal distributions are, generally, not uniform.

Ex. 3.2.26 Let $X \sim \mathrm{UNI}[0,1]$ and $Y|X = x_0 \sim \mathrm{UNI}[0,x_0]$. Then

$$f_{Y|X}(y|X = x) = \begin{cases} \frac{1}{x} & : 0 < y < x \\ 0\text{otherwise} \end{cases}$$

so that

$$f(x,y) = f_{Y|X=x}(y|X = x) \cdot f_X(x) = \begin{cases} \frac{1}{x} & : 0 < x < 1, 0 < y < x \\ 0 & : \text{otherwise} \end{cases}$$

Then

$$f_Y(y) = \int_{-\infty}^{\infty} f(x,y)\,\mathrm{d}x = \begin{cases} \int_y^1 \frac{1}{x}\,\mathrm{d}x = -\ln(y) & : 0 < y < 1 \\ 0\text{otherwise} \end{cases}$$

3.3 Expectation

Let $f(x,y)$ be the joint density function for (X,Y). Then $\mathbb{E}(g(X,Y)) = \iint g(x,y)f(x,y)\,\mathrm{d}x\,\mathrm{d}y$. Suppose $g(x,y)$ is uniformly distributed

$$\begin{aligned} \mathbb{E}(|X-Y|) &= \iint_{[0,60]^2} \frac{|x-y|}{60^2}\,\mathrm{d}x\,\mathrm{d}y \\ &= 2\int_0^{60}\int_0^x \frac{x-y}{60^2}\,\mathrm{d}y\,\mathrm{d}x \\ &= \frac{1}{60^2}\int_0^{60}\left[xy - \frac{y^2}{2}\right]_{y=0}^{y=x}\mathrm{d}x \\ &= \frac{2}{60^2}\int_0^{60}\frac{x^2}{2}\,\mathrm{d}x \\ &= 20 \end{aligned}$$

As an application, suppose X is hypergeometric where $\mathbb{P}(X=k) = \frac{\binom{M}{k}\cdot\binom{N-M}{n-k}}{\binom{N}{n}}$.

Ex. 3.3.1 Suppose there are N hunters waiting for a flock of K ducks to fly by. Each hunter picks a duck at ranom and hits it with probability p. Let X count the number of ducks which survive. What is $\mathbb{E}(X)$?

Let $\eta_i = 1$ if duck i survives, and 0 otherwise. Then $\mathbb{E}(\eta_i) = \left(1-\frac{p}{K}\right)^N$ for all i. Thus $\mathbb{E}(X) = K\,\mathbb{E}(\eta_i)$.

Ex. 3.3.2 Suppose $X \sim \text{Neg.Binom}(r,p)$. Then $\mathbb{P}(X=k) = \binom{k-1}{r-1}p^r q^{k-r}$ for each $k = r, r+1, \ldots$. Let T_j denote the number of trials in which success j happens after turn $(j-1)$, so $T_j \sim \text{Geom}(P)$. Then $\mathbb{E}(T_j) = 1/p$ and $\mathbb{E}(X) = r/p$.

Ex. 3.3.3 Suppose we have M "A" symbols and N "B" symbols, where any of the $(M+N)!$ arrangements are equally likely. Let R denote the number of runs. What is $\mathbb{E}(R)$?

Write $R = X+Y$ where X counts the number of "A" runs, and Y the number of "B" runs. Let $X = X_1 + X_2 + \cdots + X_{M+N}$ where

$$X_k = \begin{cases} 1 & \text{if an } A \text{ run starts at position } k \\ 0 & \text{otherwise} \end{cases}$$

We have

$$\mathbb{E}(X_1) = \mathbb{P}(\text{first symbol } A) = \frac{M}{M+N}$$

and for $k \geq 2$

$$\mathbb{E}(X_k) = \mathbb{P}(\text{symbol } k-1 \text{ is } B, \text{ symbol } k \text{ is } A) = \frac{N\cdot m}{(M+N)(M+N-1)}$$

Thus

$$\begin{aligned}\mathbb{E}(R) &= \mathbb{E}(X) + \mathbb{E}(Y) \\ &= \sum_{k=1}^{M+N}\mathbb{E}(X_k) + \sum_{k=1}^{M+N}\mathbb{E}(Y_k) \\ &= \mathbb{E}(X_1) + (M+N-1)\,\mathbb{E}(X_k) + \mathbb{E}(Y_1) + (M+N-1)\,\mathbb{E}(Y_k) \\ &= \frac{M}{M+N} + \frac{MN}{M+N} + \frac{N}{M+N} + \frac{MN}{M+N} \\ &= 1 + \frac{2MN}{M+N}\end{aligned}$$

Def'n. 3.3.4 *For two variables X and Y, let* $\text{Cov}(X,Y) = \mathbb{E}(X\cdot Y) - \mathbb{E}(X)\cdot\mathbb{E}(Y) = \mathbb{E}((X-\mathbb{E}(X))(Y-\mathbb{E}(Y)))$.

If X and Y are independent, then $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$, so $\text{Cov}(X,Y) = 0$.

Ex. 3.3.5 Let X be -1, 0, or 1 with probability 1/3 each, and $Y = X^2$. Then $\mathbb{E}(X) = 0$ and $\mathbb{E}(XY) = \mathbb{E}(X) = 0$, so $\text{Cov}(X,Y) = 0$.

Prop. 3.3.6 *1.* $\mathrm{Cov}(X,X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \mathrm{Var}(X) \geq 0$ *is almost surely positive definite.*

2. $\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$

3. Given variables X_1, X_2, Y *and constans* $a_1, a_2, c \in \mathbb{R}$,

$$\mathrm{Cov}(a_1X_1 + a_2X_2, Y) = a_1\mathrm{Cov}(X_1,Y) + a_2\mathrm{Cov}(X_2,Y)$$

so the covariance is bilinear.

4. $|\mathrm{Cov}(X,Y)| \leq \sigma(X)\sigma(Y)$. *We often define* $\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma(X)\sigma(Y)}$ *is the correlation coefficient of* X *and* Y. *Note that* $\rho(aX+c, bY+d) = \frac{ab\mathrm{Cov}(X,Y)}{|a||b|\sigma(X)\sigma(Y)} = \rho(X,Y)\,\mathrm{sgn}(a)\,\mathrm{sgn}(b)$.

Ex. 3.3.7 Suppose N phones are distributed randomly among N people, where X is the number of matches. Write $X = \eta_1 + \eta_2 + \cdots + \eta_N$, where η_k is 1 if person k is matched with her phone, and 0 otherwise. Then $\mathbb{E}(X) + N\,\mathbb{E}(\eta_1) = 1$. Similarly,

$$\begin{aligned}\mathrm{Var}(X) &= \mathbb{E}(X^2) - \mathbb{E}(X)^2\\ &= 2 - 2\sum_{i<j}\mathbb{E}(\eta_i\eta_j)\\ &= 2 - 2\binom{N}{2}\frac{1}{N(N-1)}\\ &= 1\end{aligned}$$

since

$$\mathbb{E}(\eta_i\eta_j) = \frac{1}{N}\mathbb{P}(\eta_j = 1|\eta_i = 1) = \frac{1}{N}\cdot\frac{1}{N-1}$$

Lemma 3.3.8 $\rho(X,Y) \in [-1,1]$ *and* $\rho(X,Y) = 1$ *iff* $Y = mX + b$ *for some* $m > 0$ *(and* -1 *for some* $m < 0$*).*

Proof Let σ_X, σ_Y denote the stanard deviations, and $Z = \frac{X}{\sigma_X} + \frac{Y}{\sigma_Y}$. We then have

$$\begin{aligned}0 \leq \mathrm{Var}(Z) = \mathrm{Var}\left(\frac{X}{\sigma_X} + \frac{Y}{\sigma_Y}\right)\\ &= \mathrm{Var}\left(\frac{X}{\sigma_X}\right) + \mathrm{Var}\left(\frac{Y}{\sigma_Y}\right) + 2\mathrm{Cov}\left(\frac{X}{\sigma_X} + \frac{Y}{\sigma_Y}\right)\\ &= \frac{\mathrm{Var}\,X}{\sigma_X^2} + \frac{\mathrm{Var}\,Y}{\sigma_Y^2} + 2\frac{\mathrm{Cov}(X,Y)}{\sigma_X\sigma_Y}\\ &= 2(1 + \rho(X,Y))\end{aligned}$$

In other words that $1 + \delta(X,Y) \geq 0$ if and only $\rho(X,Y) \geq -1$, while $\mathrm{Var}(Z) = 0$ if and only if $\rho(X,Y) = -1$. Then $\frac{X}{\sigma_X} + \frac{Y}{\sigma_Y} = d$ almost surely, so $Y = mX + b$ for $m < 0$. The same argument works for $W = \frac{X}{\sigma_X} - \frac{Y}{\sigma_Y}$ in the positive case. □

Ex. 3.3.9 Consider a process with two states and 3 paths. From state 0, we return to state 0 in 5 minutes with probability 1/3, in 7 minutes with probability 1/3, and travel to state 1 in 3 minutes with probability 1/3. What is the expected amount of time to go to state 1 from state 0?

Let X be a random variable denoting the choices, so $X = 1$ if we choose A, 2 if we choose B, or 3 if we choose C. Then

$$T|X = \begin{cases} 3 & X = 1 \\ 5 + T & X = 2 \\ 7 + T & X = 3 \end{cases}$$

so that

$$\mathbb{E}(T|X) = \begin{cases} 3 & X = 1 \\ 5 + \mathbb{E}(T) & X = 2 \\ 7 + \mathbb{E}(T) & X = 3 \end{cases}$$

and

$$\mathbb{E}(T) = \mathbb{E}(\mathbb{E}(T|X)) = \frac{1}{3} \cdot 3 + \frac{1}{3}(5 + \mathbb{E}(T)) + \frac{1}{3}(7 + \mathbb{E}(T))$$

Suppose $g(x, y)$ is an arbitrary function from $\mathbb{R}^2 \to \mathbb{R}$ such that the first moments below are integrable. Then $\mathbb{E}(g(X, Y)) = \mathbb{E}(\mathbb{E}(g(X, Y)|X))$. Note that if $g(X, Y) = h_1(X) \cdot h_2(Y)$ (i.e. the variables separate), then

$$\mathbb{E}(h_1(X)h_2(Y)|X) = h_1(X) \cdot \mathbb{E}(h_2(Y)|X)$$

Ex. 3.3.10 Suppose $X \text{UNI}[0, 1]$ and $Y|X \sim \text{UNI}[0, X]$. Then

$$\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

Note that $\mathbb{E}(X) = 1/2$ and $\mathbb{E}(Y|X) = X/2$. We can compute

$$\mathbb{E}(Y) = \mathbb{E}(\mathbb{E}(Y|X)) = \mathbb{E}(X/2) = \frac{1}{2}\mathbb{E}(X) = \frac{1}{4}$$

and

$$\mathbb{E}(XY) = \mathbb{E}(\mathbb{E}(XY|X)) = \mathbb{E}(X\,\mathbb{E}(Y|X)) = \frac{1}{2}\mathbb{E}(X^2) = \frac{1}{6}$$

Ex. 3.3.11 (Sum with random number of terms) Suppose that the number of customers which enter a store with an hour is a random variable N with $\mathbb{E}(N) = n$ fixed. Suppose each customer k spends a random amount X_k with $X_k) = \mu$ fixed. We assume that the variables $X_1, X_2, \ldots, X_k$ and N are independent. Let Y denote the income within the hour, so $Y = X_1 + X_2 + \cdots + X_N$. We then have

$$\mathbb{E}(Y) = \mathbb{E}(\mathbb{E}(Y|N)) = \mathbb{E}(N\mu) = \mu\,\mathbb{E}(N) = n \cdot \mu$$

Let's compute the conditional variance.

$$\text{Var}(X|Y) = \mathbb{E}(X^2|Y) - (\mathbb{E}(X|Y))^2$$

Taking the expected value, we have

$$\mathbb{E}(\mathrm{Var}(X|Y)) = \mathbb{E}(\mathbb{E}(X^2|Y)) - \mathbb{E}((\mathbb{E}(X|Y))^2)$$

as well, since $\mathbb{E}(X|Y)$ is a random variable, we have

$$\mathrm{Var}(\mathbb{E}(X|Y)) = \mathbb{E}((\mathbb{E}(X|Y))^2) - (\mathbb{E}(\mathbb{E}(X|Y)))^2$$

so that

$$\begin{aligned}\mathbb{E}(\mathrm{Var}(X|Y)) + \mathrm{Var}(\mathbb{E}(X|Y)) &= \mathbb{E}(\mathbb{E}(X^2|Y)) - [\mathbb{E}(\mathbb{E}(X|Y))]^2\\ &= \mathbb{E}(X^2) - [\mathbb{E}X]^2\\ &= \mathrm{Var}(X)\end{aligned}$$

Ex. 3.3.12 Suppose the departure time of a train is $T \sim \mathrm{UNI}[0,t]$. Passengers board the train according to a Poisson process with intensity λ. Let X denote the number of passengers who take the train. What is $\mathbb{E}(X)$, $\mathrm{Var}(X)$?

We have

$$\begin{aligned}\mathbb{E}(X) &= \mathbb{E}(\mathbb{E}(X|T)) = \mathbb{E}(\lambda T)\\ &= \lambda\,\mathbb{E}(T) = \frac{\lambda t}{2}\end{aligned}$$

since to compute $\mathbb{E}(X|T=s)$, recall that $X|T=s \sim \mathrm{Poi}(\lambda s)$, so $\mathbb{E}(X|T=s) = \lambda s$. As well,

$$\begin{aligned}\mathrm{Var}(X) &= \mathbb{E}(\mathrm{Var}(X|T)) + \mathrm{Var}(\mathbb{E}(X|T))\\ &= \lambda\,\mathbb{E}(T) +^2 \mathrm{Var}(T)\\ &= \frac{\lambda t}{2} + \frac{\lambda^2 t^2}{12}\end{aligned}$$

Consider one random variable X. How can we replace it with something determinstic?

Suppose we choose the value $c \in \mathbb{R}$ for which $\mu = \mathbb{E}(X)$. Then

$$\begin{aligned}d(0) &= \mathbb{E}((c-X)^2)\\ &= \mathbb{E}((X-\mu)^2 + (\mu-c)^2 + 2(X-\mu)(\mu-c))\\ &= \mathrm{Var}(X) + (\mu-c)^2\end{aligned}$$

which is minimized by $c=\mu$; in other words, μ minimizes the mean square displacement from c, and the mean square displacement is the variance. If instead we minimize the mean difference, we end up with the median.

Ex. 3.3.13 Suppose $S\mathcal{N}(\mu,\sigma^2)$, and we observe $R|S \sim \mathcal{N}(S,1)$, in other words the signal S with some random noise. If a value r is received, what is our guess for the value sent? We want $\mathbb{E}(S|R=r)$.

We want to find

$$f_{S|R}(s|R=r) = \frac{f(s,r)}{f_R(r)} = \frac{f_{R|S}(r|S=s)f_S(s)}{f_R(r)} =: V$$

3.4 Moment Generating Functions

Def'n. 3.4.1 *Given a random variable X, the **moment generating function** of X is*

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

Note that such a function may not be finite!

We have some properties of moment generating functions.

1. $M_X(0) = \mathbb{E}(1) = 1$
2. $M_X'(t) = \mathbb{E}(Xe^{tX})$ so $M_X'(0) = \mathbb{E}(X)$.
3. In general, $M_X^{(n)}(t) = \mathbb{E}(X^n e^{tX})$ so $M_X^{(n)}(0) = \mathbb{E}(X^n)$.

Ex. 3.4.2 (Common Moment Generating Functions) 1. Suppose $X \sim \operatorname{Binom}(n,p)$. Then

$$\begin{aligned} M_X(t) = \mathbb{E}(e^{tX}) &= \sum_{k=0}^{n} e^{tk}\binom{n}{k}p^k(1 = p)^{n-k} \\ &= \sum_{k=0}^{n}\binom{n}{k}(pe^t)^k(1-p)^{n-k} \\ &= (pe^t + 1 - p)^n \end{aligned}$$

2. Suppose $X \sim \operatorname{Poi}(\lambda)$. Then

$$\begin{aligned} M_X(t) &= \mathbb{E}(e^t X) \\ &= \sum_{k=0}^{\infty} e^{tk}\frac{\lambda^k}{k!}e^{-\lambda} \\ &= e^{-\lambda}\sum_{k=0}^{\infty}\frac{(\lambda e^t)^k}{k!} \\ &= e^{-\lambda}e^{\lambda e^t} \\ &= e^{\lambda(e^t-1)} \end{aligned}$$

3. Suppose $Z \sim \mathcal{N}(0,1)$. Then

$$\begin{aligned} M_Z(t) = \mathbb{E}(e^{tZ}) &= \int_{-\infty}^{\infty} e^{tx}\phi(x)\,\mathrm{d}x \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{tx}e^{-x^2/2}\,\mathrm{d}x \\ &= \frac{1}{\sqrt{2\pi}}e^{t^2/2}\int_{-\infty}^{\infty} e^{-(x-t)^2/2}\,\mathrm{d}x \\ &= e^{t^2/2} \end{aligned}$$

and if $X = \sigma Z + \mu$,

$$M_X(t) = \mathbb{E}(e^{tX}) = e^{t\mu}\,\mathbb{E}(e^{t\sigma Z}) = e^{t\mu}M_Z(t\sigma)$$

Suppose X and Y are indpendent. Then their moment generating functions are also independent. This follows since

$$M_{X+Y}(t) = \mathbb{E}(e^{t(X+Y)}) = \mathbb{E}(e^{tX}e^{tY}) = \mathbb{E}(e^{tX})\,\mathbb{E}(e^{tY}) =_X (t)M_Y(t)$$

Given two random variables X_1 and X_2, their joint moment generating function is

$$M : \mathbb{R}^2 \to \mathbb{R} \quad M(t_1,t_2) = \mathbb{E}(e^{(t_1X_1+t_2X_2)})$$

3.5 Multivariate Normal Distribution

Def'n. 3.5.1 *Given random variables $(X_1, X_2, \dots, X_n)$ random variables, define the matrix C with $C_{ij} = \operatorname{Cov}(X_i, X_j)$. This is a symmetric and positive definite matrix. To se that it is positive definition, let $a \in \mathbb{R}^n$ be arbitrary, so that*

$$\begin{aligned} a^TCa &= \sum_{i=1}^n a_i(ca)_i \\ &= \sum_{i=1}^n \sum_{j=1}^n a_i c_{ij} a_j \\ &= \sum_{i,j=1}^n a_i a_j \operatorname{Cov}(X_i X_j) \\ &= \operatorname{Cov}\left(\sum_{i=1}^n a_i X_i, \sum_{j=1}^n a_j X_j\right) \\ &= \operatorname{Var}\left(\sum_{i=1}^n a_i X_i\right) > 0 \end{aligned}$$

so as long as the random variables are linearly independent, then the covariance matrix is positive definitite.

Suppose (X_1, X_2) have joint density $f(x_1, x_2)$. If $Y = (Y_1, Y_2) = k(X)$ where k is an injective smooth map, the joint density $g(y_1, y_2)$ is given by

$$g(y_1,y_2)\frac{1}{\det(J(k))}f(k^{-1}(y_1,y_2))$$

Def'n. 3.5.2 *Given $\mu = (\mu_1, \mu_2, \dots, \mu_n) \in \mathbb{R}^n$, we say that the random variables $(X_1, X_2, \dots, X_n)$ are jointly normally distributed if their joint density satisfies*

$$f(x_1,x_2,\dots,x_n) = \frac{\sqrt{\det C^{-1}}}{(2\pi)^{n/2}}\exp\left(-\frac{1}{2}(x-\mu)^TA(x-\mu)\right)$$

where C is the covariance matrix.

When $\mu = 0$ and $C = I_n$, then

$$f(x) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left(-\frac{1}{2}x_1^2\right)\exp\left(-\frac{1}{2}x_2^2\right)\cdots\exp\left(-\frac{1}{2}x_n^2\right)$$

in other words that $f(x1,\dots,x_n) = f(x_1)f(x_2)\cdots f(x_n)$.

Prop. 3.5.3 *If $(X_1,\dots,X_n)$ are jointly normal, then there exists an invertible matrix B so that $X = By + \mu$ where $Y = (Y_1, Y_2, \dots, Y_n)$ are as above.*

Proof Since A is symmetric and positive definite, let $\lambda_1, \lambda_2, \dots, \lambda_n$ be the positive eigenvalues and (u_i) be the associated eigenvectors. Write $P = (u_1 \cdots u_n)$ is a diagonalizing matrix and $A = PDP^{-1}$. Define $B^{-1} = D^{1/2}P^{-1}$ so that $A = (B^{-1})^T B^{-1}$ and $B = PD^{-1/2}$. Thus $BB^T = PD^{-1/2}D^{-1/2}P^{-1} = A^{-1} + C$. We also want $X = BY + \mu$ so $Y = B^{-1}(X - \mu) = k(X)$ so $J = \det(B^{-1}) = \frac{1}{\det B}$.

Now, let's compute the joint density of Y. Note that $\det(A) = \det((B^{-1})^T B^{-1}) = (\det B^{-1})^2$ so $\sqrt{\det A} = \frac{1}{\det B}$. As well, $(x-\mu)^T A(x-\mu) = (x-\mu)^T (B^{-1})^T B^{-1}(x-\mu) = y^T y$. We have

$$\begin{aligned} g(y) &= \frac{1}{J} \cdot f(x) \\ &= \det B \cdot \frac{\sqrt{a}}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}(x-\mu)^T A(x-\mu)\right) \\ &= \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}(y_1^2 + \cdots + y_n^2)\right) \end{aligned}$$

as required.

What is the covariance matrix of $(X_1, X_2, \dots, X_n)$? We have

$$\begin{aligned} \mathrm{Cov}(X_i, X_j) &= \mathrm{Cov}\left(\sum_{k=1}^n B_{ik} Y_k + \mu_i, \sum_{l=1}^n B_{jl} Y_l + \mu_j\right) \\ &= \sum_{k,l=1}^n B_{ik} B_{jl} \mathrm{Cov}(Y_k, Y_l) \\ &= \sum_{l=1}^n B_{il} B_{jl} \\ &= \sum_{l=1}^n B_{il} (B^T)_{lj} \\ &= (B \cdot B^T)_{ij} \\ &= C_{ij} \end{aligned}$$

In particular, if $C_{ij} = 0$ for $i \neq j$, then A_{ij} is diagonal so $f(x)$ is a product and $X_1, \dots, X_n$ are independent. Thus for jointly normally distributed variables, uncorrelated variables are independent. □

When $n = 2$, we have

$$C = \begin{pmatrix} \sigma_1^2 & \delta\sigma_1\sigma_2 \\ \delta\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}$$

Thus $\det C = (1 - \delta^2)\sigma_1^2\sigma_2^2$, *so*

$$A = C^{-1} = \frac{1}{1-\delta^2} \begin{pmatrix} \frac{1}{\sigma_1^2} & -\frac{\delta}{\sigma_1\sigma_2} \\ -\frac{\delta}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2} \end{pmatrix}$$

Now if $x = (x_1, x_2)$, then

$$(x-\mu)^T A(x-\mu) = \frac{1}{1-\delta^2}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2} - 2\delta\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}\right)$$

As well,

$$f_{X_1|X_2}(x_1|X_2 = x_2) = \frac{f(x_1,x_2)}{f_{X_2}(x_2)} = C(x_2)\exp\left(-\frac{1}{2}\cdot\frac{1}{(1-\delta^2)\sigma_1^2}(x_1^2 - 2x_1\mu(x_2))\right)$$

where $\mu(x_2) = \mu_1 + \delta\frac{\sigma_1}{\sigma_2}(x_2-\mu_2)$. Thus $X_1|X_2 = x_2 \sim \mathcal{N}(\mu(x_2),(1-\delta^2)\sigma_1^2)$.

3.6 Common Inequalities

Prop. 3.6.1 (Markov's Inequality) *Let X be a non-negative random variable, i.e. $X \geq 0$ almost surely. Fix $\mu = \mathbb{E}(X)$; then, for any $x > 0$, $\mathbb{P}(X \geq a) \leq \mu/a$.*

Proof Let η_a be the indicator for the event $X \geq a$. Then $\mathbb{E}(\eta_a) = \mathbb{P}(X \geq a)$, while $a \cdot \eta_a \leq X$. Then $a \cdot \mathbb{E}(\eta_a) \leq \mathbb{E}(X)$ and we are done. □

Prop. 3.6.2 (Chebyshev's Inequality) *Let X be an arbitrary random variable with $\mu = \mathbb{E}(X)$ and $\sigma^2 = \mathrm{Var}(X)$. Then, for any $b > 0$, $\mathbb{P}(|X-\mu| \geq b) \leq \frac{\sigma^2}{b^2}$.*

Proof Fix $Y = (X-\mu)^2$, so $Y \geq 0$ and $\mathbb{E}(Y) = \sigma^2$. Then by Markov, $\mathbb{P}((X-\mu) \geq b) = \mathbb{P}(Y \geq b^2) \leq \mathbb{E}(Y)/b^2 = \sigma^2/b^2$. □

Prop. 3.6.3 (Weak Law of Large Numbers) *Let $X_1, X_2, \ldots, X_n$ be an i.i.d. sequence of random variables with $\mu = \mathbb{E}(X_k)$ and $\sigma^2 = \mathrm{Var}(X_k)$. Then for any $\epsilon > 0$, $\mathbb{P}\left(|\frac{S_n}{n} - \mu| \geq \epsilon\right) \to 0$ as $n \to \infty$.*

Proof Note that $\mathbb{E}(S_n) = n\mu$ and $\mathrm{Var}(S_n) = n\sigma^2$ by independence. Then

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| \geq \epsilon\right) = \mathbb{P}(|S_n - n\mu| \geq n\epsilon) \leq \frac{\mathrm{Var}(S_n)}{(n\epsilon)^2} = \frac{1}{n}\cdot\frac{\sigma^2}{\epsilon^2}$$

which tends to 0 as n goes to infinity. □

As a special case, let X_k be Bernoulli with probability $p \in (0,1)$, and we get Bernoulli's law of large numbers.

Now, consider $(S_n - n\mu)/n^\alpha$. Then

$$\mathbb{P}\left(\left|\frac{S_n - n\mu}{n^\alpha}\right| \geq \epsilon\right) = n^{1-2\alpha}\frac{\sigma^2}{\epsilon^2} \to 0$$

whenever $1/2 < \alpha$. This suggests that the order of the fluctuations is actually $\sqrt{n}$.

Prop. 3.6.4 (Central Limit Theorem) *Let X_1, X_2, be iid random variables with $\mathbb{E}(X_1) = \mu$ and* $\mathrm{Var}(X_1) = \sigma^2$. *Then for any $a \in \mathbb{R}$,*

$$\mathbb{P}\left(\frac{S_n - n\mu}{\sqrt{n}\cdot\sigma} \leq a) \to \Phi(a)$$

as n goes to infinity.

We will prove the Central Limit Theorem under the additional assumption that $M(t) = \mathbb{E}(e^{tX_k})$ exists for some $t_0 > 0$. We also have the following lemma (without proof):

Lemma 3.6.5 *Let $Y_1, Y_2, \ldots$ be a sequence of random variables with moment generation functions, and if W is another random variable. Let $M_n(t)$ be the moment generating function for Y_n, and $M_W(t)$ for W. Then of $M_n(t) \to M_W(t)$ for all $t \in \mathbb{R}$, then $F_n(x) \to F_W(x)$.*

We now prove the Central Limit Theorem

Proof First suppose $\mu = 0$ and $\sigma^2 = 1$. Then

$$\frac{S_n - n\mu}{\sqrt{n}\sigma} = \frac{S_n}{\sqrt{n}}$$

Define $\Psi(t) = \log M(t)$. Then $\Psi(0) = 0$, $\Psi'(0) = \mu = 0$, and $\Psi''(t) = \mathrm{Var}\, x = \sigma^2 = 1$. We want to show that

$$M_{\frac{S_n}{\sqrt{n}}}(t) {\to} M_Z(t) {=} e^{t^2/2} {\Leftrightarrow} \Psi_{\frac{S_n}{\sqrt{n}}}(t) {\to} \Phi_Z(t) {=} \frac{t^2}{2}$$

First note that

$$M_{\frac{S_n}{\sqrt{n}}}(t) = \mathbb{E}\left(e^{t\frac{S_n}{\sqrt{n}}}\right) = M_{S_n}\left(\frac{t}{\sqrt{n}}\right)$$

and by independence,

$$M_{S_n(t)} = M_{X_1}(t) M_{X_2}(t) \cdots M_{X_n}(t) = (M(t))^n$$

Now $\Psi(s) = \Psi(0) + \Psi'(0)s + \Psi''(0)(0)\frac{s^2}{2} + O(s^3) = \frac{s^2}{2} + O(s^3)$ so that

$$\Psi_{\frac{S_n}{\sqrt{n}}}(t) = n\Psi\left(\frac{t}{\sqrt{n}}\right) = \frac{t^2}{2} + O\left(\frac{t^3}{\sqrt{n}}\right)$$

and as $n \to \infty$, this tends to $t^2/2$.

In general, let $\mu \in \mathbb{R}$, $\sigma^2 > 0$ be arbitrary. Then $X_k^* = \frac{X_k - \mu}{\sigma}$ is an iid sequence and

$$\frac{S_n - n\mu}{\sqrt{n}\sigma} = \frac{S_n^*}{\sqrt{n}}$$

and the previous statement applies. □

Ex. 3.6.6 Let $X_1, X_2, \ldots$ be independent measurements of a quantity μ. Determine $\delta(n)$ so that $\mathbb{P}\left(\left|\frac{S_n}{n} - \mu\right| \leq \delta(n)\right) \geq 0.99$.

We want

$$\begin{aligned} 0.99 &\leq \mathbb{P}(|S_n - n\mu| \leq n\delta) \\ &= \mathbb{P}\left(||\frac{S_n - n\mu}{\sqrt{n}\sigma}| \leq \frac{\sqrt{n}\delta}{\sigma}\right) \\ &= 2\Phi\left(\frac{\sqrt{n}\delta}{\sigma}\right) - 1 \end{aligned}$$

We usually do not have σ! Instead, let

$$\overline{X} = \frac{S_n}{n}, \frac{(X_1 - \overline{X})^2 + \cdots + (X_n - \overline{X})^2}{n-1} = s_n^2$$

To justify the $n-1$ in the denominator, note that Note that $\mathbb{E}[(X_1 - X)^2] = \mathrm{Var}(X_1) + \mathrm{Var}(\overline{X}) - 2\,\mathrm{Cov}(X_1, \overline{X})$. Then $\mathrm{Var}(X_1) = \sigma^2$ so $\mathrm{Var}(\overline{X}) = \frac{\sigma^2}{n}$, while $\mathrm{Cov}(X_1, \overline{X}) = \mathrm{Cov}(X_1, S_n/n) = \frac{1}{n}\sum_{k=1}^{n} \mathrm{Cov}(X_1, X_k) = \mathrm{Var}(X_1)/n$. Now

$$\begin{aligned} \mathbb{E}((X_1 - \overline{X})^2 + \cdots + (X_n - \overline{X})^2) &= \mathbb{E}[(X_1 - X)^2] + \cdots + \mathbb{E}[(X_n - X)^2] \\ &= n\left(\sigma^2 - \frac{\sigma^2}{n}\right) \\ &= (n-1)\sigma^2 \end{aligned}$$

so that $\mathbb{E}(s_n^2) = \sigma^2$.

Get 40 random real number which is a fixed integer value plus a uniform error on $[-1/2, 1/2]$. They are added up and rounded up to the closest integer T_1. Let T_2 be the sum of the individual errors. What is $\mathbb{P}(T_1 = T_2)$?