Honours Probability - MATH 458

Professor Chen - lnchen@math.mcgill.ca

Office Hours: TBA - Burn 1230

Contents

1 The Basics of Probability

1.1 Overview of a Probability Space

A probability space $(\Omega, \mathcal{S}, \mathbb{P})$ is a model of random trials/experiments. This means

- all possible outcomes are known: e.g. flipping a coin.
- the outcome of any specific performance is unknown
- repeatable under identical conditions

Definition 1.1. $\Omega = \{\text{all possible outcomes}\}$ is the sample space. $\omega \in \Omega$ is called a sample point.

Definition 1.2. $\mathcal{S} = \{\text{all the events we can study}\}$ is a σ-field. $\mathcal{S}$ is a collection of subsets of Ω. $A \in \mathcal{S}$ is an event. $\emptyset, \Omega$ are the trivial events.

Definition 1.3. $\mathbb{P} : A \in \mathcal{S} \mapsto [0,1]$ with the requirement that $\mathbb{P}(\emptyset) = 0, \mathbb{P}(\Omega) = 1$ is the probability.

Example 1.4. Flip a coin once: $\Omega = \{H, T\}; \mathcal{S} = \{\{H\}, \{T\}, \{H, T\}, \emptyset\} = 2^{\Omega} := \{\text{all subsets of } \Omega\}$.

$$\mathbb{P}(\{H\}) = \frac{1}{2} = \mathbb{P}(\{T\})$$

Example 1.5. Roll a fair die once: $\Omega = \{1,2,3,4,5,6\}; \mathcal{S} = 2^{\Omega}$.

$$\mathbb{P}(\{1\}) = \frac{1}{6}, \mathbb{P}(\{2,3\}) = \frac{1}{3}$$

Example 1.6. Choose any number at random from $[0,1]$. $\Omega = [0,1]$.

$$\mathbb{P}\left(\left\{\frac{1}{2}\right\}\right) = 0, \mathbb{P}\left(\left[0, \frac{1}{2}\right]\right) = \frac{1}{2}$$

Example 1.7. Consider $\Omega = \mathbb{N}$, $\mathcal{S} = 2^{\mathbb{N}}$. $\mathbb{P}(n) = 2^{-n}$.

Definition 1.8. $\mathcal{S}$ is a collection of subsets of Ω. $\mathcal{S}$ is called a σ-field if

- $\Omega \in \mathcal{S}$
- if $A \in \mathcal{S}$ then $A^c \in \mathcal{S}$

- if $A_i \in \mathcal{S}\ \forall i \in \mathbb{N}$ then $\cup_i A_i \in \mathcal{S}$

Some results of this are

- $\emptyset \in \mathcal{S}$
- if $A_i \in \mathcal{S}, i \in \{1, ..., N\}$ then $\cup_{i=1}^{N} A_i \in \mathcal{S}$, $\cap_{i=1}^{N} A_i \in \mathcal{S}$
- if $A_i \in \mathcal{S}\ \forall i \in \mathbb{N}$ then $\cap_i A_i \in \mathcal{S}$
- if $A, B \in \mathcal{S}$ then $B \backslash A \in \mathcal{S}$

Definition 1.9. Let C be a collection of subsets of Ω. The σ-field generated by C, denoted $\sigma(C)$ is the smallest σ-field that contains C, i.e., if $\mathcal{S}$ is a σ-field and $C \subset \mathcal{S}$ then $\sigma(C) \subset \mathcal{S}$.

Remark 1.10. For $A \in \mathcal{S}$ the smallest σ-field containing A, i.e., the σ-field generated by A is $\mathcal{S} = \{\emptyset, \Omega, A, A^c\}$.

Remark 1.11. For $A, B \in \mathcal{S}$ the smallest σ-field containing A, B, i.e., the σ-field generated by A, B is $\mathcal{S} = \{\emptyset, \Omega, A, A^c, B, B^c, A \cup B, A \cap B, A \cup B^c, A^c \cup B, (A \cup B)^c, (A \cup B)^c \cup (A \cap B)\}$.

Proposition 1.12.

$$\sigma(C) = \cap\{\mathcal{S} : \mathcal{S} \textit{ is a } \sigma\textit{-field and } C \subset \mathcal{S}\}$$

Note that any intersection of σ-fields is a σ-field.

Proof. By definition of $\sigma(C)$, $\sigma(C)$ is an element of the $\{\mathcal{S} : \mathcal{S}$ is a σ-field and $C \subset \mathcal{S}\}$, so we have the inclusion $\supset$.
We also have that the right hand side is a σ-field that contains C so we also have the opposite inclusion, since $\sigma(C)$ is the smallest of these. $\square$

Remark 1.13. If $\mathcal{S}_1, \mathcal{S}_2$ are σ-fields then $\mathcal{S}_1 \cup \mathcal{S}_2$ is not necessarily a σ-field.

Definition 1.14. The Borel σ-field on $\mathbb{R}$, denoted by $\mathcal{B}(\mathbb{R})$, is the σ-field generated by all the open intervals on $\mathbb{R}$, i.e., $\mathcal{B} = \sigma(\{(a, b) : a < b; a, b \in \mathbb{R}\}$.

Remark 1.15.

- $\forall a \in \mathbb{R} : (a, \infty), (-\infty, a) \in \mathcal{B}(\mathbb{R})$
- $\forall a \in \mathbb{R} : \{a\} \in \mathcal{B}(\mathbb{R})$
- Any interval is in $\mathcal{B}(\mathbb{R})$

Remark 1.16. There exist subsets of $\mathbb{R}$ that are not Borel.

Definition 1.17. Let Ω be a sample space and $\mathcal{S}$ a σ-field on Ω. A set function $\mathbb{P}$ defined on $\mathcal{S}$ is a probability (probability measure) if

- $\forall A \in \mathcal{S}, \mathbb{P}(A) \geq 0$
- $\mathbb{P}(\Omega) = 1$

- It has countable additivity, i.e., if $A_i \in \mathcal{S}; i \in \mathbb{N}$; and $\forall i \neq j, A_i \cap A_j = \emptyset$ then $\mathbb{P}\left(\cup_i A_i\right) = \sum_i \mathbb{P}(A_i)$

Remark 1.18. Some properties of $\mathbb{P}$ are

- $\mathbb{P}(\emptyset) = 0$. $\mathbb{P}(\emptyset) = \sum_{i=1}^{\infty} \mathbb{P}(\emptyset) \implies \mathbb{P}(\emptyset) = 0$.
- Finite Additivity
- $\forall A \in \mathcal{S}, \mathbb{P}(A^c) = 1 - \mathbb{P}(A)$
- if $A, B \in \mathcal{S}$ and $A \subset B$ then $\mathbb{P}(B - A) = \mathbb{P}(B) - \mathbb{P}(A)$. Note that $\mathbb{P}(A) < \mathbb{P}(B)$.
- if $A, B \in \mathcal{S}$ then $\mathbb{P}(B \cup A) = \mathbb{P}(B) + \mathbb{P}(A) - \mathbb{P}(A \cap B)$. This is the addition rule. Note that this implies finite subadditivity, i.e., $\mathbb{P}(\cup_i A_i) \leq \sum_i \mathbb{P}(A_i)$
- (Inclusion-Exclusion Formula) if $A_i \in \mathcal{S}, i \in \{1, ..., N\}$ then

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

$$= \sum_{i=1}^{N} \mathbb{P}(A_i) - \sum_{1 \leq i_1 < i_2 \leq N} \mathbb{P}(A_{i_1} \cap A_{i_2}) + ... + (-1)^{N+1}\mathbb{P}\left(\bigcap_{i=1}^{N} A_i\right)$$

$$= \sum_{k=1}^{N} (-1)^{i+1} \sum_{1 \leq i_1 < ... < i_k \leq N} \mathbb{P}\left(\bigcap_{n=i_1}^{i_k} A_n\right)$$

- (Bonferroni's Inequality) if $A_i \in \mathcal{S}, i \in \{1, ..., N\}$ then

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

Proof. (Inclusion-Exclusion Formula) For the base case of 1 set the statement is just $\mathbb{P}(A) = \mathbb{P}(A)$. Suppose it is true for $1, ..., n$ sets. Then by the addition rule

$$\mathbb{P}\left(\bigcup_{i=1}^{n+1} A_i\right) = \mathbb{P}(A_{n+1}) + \mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) - \mathbb{P}\left(\bigcup_{i=1}^{n} (A_i \cap A_{n+1})\right)$$

Applying the on the second and third terms assumption the the formula holds for n sets and then collecting terms we finally have

$$= \sum_{k=1}^{n+1} (-1)^{i+1} \sum_{1 \leq i_1 < ... < i_k \leq n+1} \mathbb{P}\left(\bigcap_{j=i_1}^{i_k} A_j\right)$$

$\square$

Proof. (Bonferroni's Inequality)
The right inequality is obvious. The statement is obvious for the base case of 1 set. Now assume it holds for N

$$\mathbb{P}\left(\bigcup_{i=1}^{N+1} A_i\right) = \mathbb{P}\left(\bigcup_{i=1}^{N} A_i \cup A_{N+1}\right) = \mathbb{P}\left(\bigcup_{i=1}^{N} A_i\right) + \mathbb{P}(A_{N+1}) - \mathbb{P}\left(\left(\bigcup_{i=1}^{N} A_i\right) \cap A_{N+1}\right)$$

$$\leq \sum_{i=1}^{N+1} \mathbb{P}(A_i) - \sum_{i_1<i_2}^{N} \mathbb{P}(A_{i_1} \cap A_{i_2}) - \sum_{i=1}^{N} \mathbb{P}(A_i \cap A_{N+1}) = \sum_{i=1}^{N+1} \mathbb{P}(A_i) - \sum_{i_1<i_2}^{N+1} \mathbb{P}(A_{i_1} \cap A_{i_2})$$

□

Remark 1.19. Ω is a sample space, $\omega \in \Omega$. $\{\omega\}$ is a singleton. If $\{\omega\} \in \mathcal{S}$ then it is an elementary event.

- Ω is discrete and finite. $\Omega = \{\omega_1, ..., \omega_n\}$ and $\mathcal{S} = 2^{\Omega}$.

It is sufficient to assign $\mathbb{P}$ on every elementary event, i.e., $\mathbb{P}(\{\omega_j\}), j = 1, ..., n$. Take the probability of any set A to be the sum of its elementary events.
One such $\mathbb{P}$ is the uniform probability with $\mathbb{P}(\{\omega_j\}) = \frac{1}{n}$ for every j.

- $\Omega = \{\omega_1, \omega_2, ...\}$ (countably infinite), $\mathcal{S} = 2^{\Omega}$

One cannot assign a uniform probability on each elementary event. However one can assign $\mathbb{P}(\{\omega_j\}) = 2^{-j}$ or

$$\mathbb{P}(\{\omega_j\}) = \begin{cases} 1: & \text{if } j = 1 \\ 0: & \text{otherwise} \end{cases} \tag{1.1}$$

- $\Omega = [0,1], \mathcal{S} = \mathcal{B}([0,1])$

Exercise: if Ω is uncountable and $\forall \omega \in \Omega, \{\omega\} \in \mathcal{S}$ then for any probability $\mathbb{P}$ on $(\Omega, \mathcal{S})$ there are at most countable ω such that $\mathbb{P}(\{\omega\}) > 0$.

Proof. Consider the sets $A_n = \{\omega \in \mathcal{S} : \mathbb{P}(\{\omega\} > \frac{1}{n}\}$. Each set has less than n elements, otherwise $\mathbb{P}(A_n) \geq n \cdot \frac{1}{n} = 1$. $\bigcup_{n=1}^{\infty} A_n$ is the set of all events with nonzero probability. It is a countable union of finite sets and is thus countable. Therefore, there are at most countable events with nonzero probability. □

Definition 1.20. Assume that $A_i \in \mathcal{S}, i \in \mathbb{N}$. $\{A_i : i \in \mathbb{N}\}$ is a non-decreasing sequence if $A_i \subset A_{i+1}, \forall i \in \mathbb{N}$. It is non-increasing if $A_i \supset A_{i+1}, \forall i \in \mathbb{N}$. If it is non-increasing or non-decreasing then it is monotone.
If $\{A_i : i \in \mathbb{N}\}$ is monotone then we have existence of $\lim_{i\to\infty} A_i = \bigcup_{i=1}^{\infty} A_i$ or $\bigcap_{i=1}^{\infty} A_i$ for non-decreasing and non-increasing respectively.

Example *1.21.* $a \in \mathbb{R}, A_i = (a - \frac{1}{i}, a + \frac{1}{i})$ then $\{A_i : i \in \mathbb{N}\}$ is non-increasing and $\lim_{i\to\infty} A_i = \{a\}$.

Example *1.22.* $A_i = (-i, i)$ then $\{A_i : i \in \mathbb{N}\}$ is non-decreasing and $\lim_{i\to\infty} = \mathbb{R}$.

Remark *1.23.* Some more properties of $\mathbb{P}$ are

- if $\{A_i : i \in \mathbb{N}\} \subset \mathcal{S}$ is non-decreasing then $\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_{i\to\infty} \mathbb{P}(A_i) \iff \mathbb{P}(\lim_{i\to\infty} A_i) = \lim_{i\to\infty} \mathbb{P}(A_i)$. (continuity from below)
- $\{A_i : i \in \mathbb{N}\} \subset \mathcal{S}$ a non-increasing sequence then $\mathbb{P}\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_{i\to\infty} \mathbb{P}(A_i) \iff \mathbb{P}\left(\lim_{i\to\infty} A_i\right) = \lim_{i\to\infty} \mathbb{P}(A_i)$. Note: Proof is from the fact that A_i^c is a non-decreasing sequence.
- $\{A_i : i \in \mathbb{N}\} \subset \mathcal{S}$ then $\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) \le \sum_{i=1}^{\infty} \mathbb{P}(A_i)$. (countable subadditivity)

Proof. (of continuity from below)
Define $B_1 = A_i, B_i = A_i - A_{i-1}, i \ge 2$. Because $A_i \uparrow$, $\{B_i\}$ are disjoint

$$\forall n \in \mathbb{N}, \bigcup_{i=1}^{n} B_i = \bigcup_{i=1}^{n} A_i = A_n$$

We know

$$\begin{aligned}
\mathbb{P}\left(\bigcup_{i=1}^{\infty} B_i\right) &= \sum_{i=1}^{\infty} \mathbb{P}(B_i) = \lim_{n\to\infty} \sum_{i=1}^{n} \mathbb{P}(B_i) \\
&= \lim_{n\to\infty} \mathbb{P}\left(\bigcup_{i=1}^{n} B_i\right) = \lim_{n\to\infty} \mathbb{P}(A_n) \\
&\implies \mathbb{P}\left(\lim_{i\to\infty} A_i\right) = \lim_{i\to\infty} \mathbb{P}(A_i)
\end{aligned}$$

□

Proof. (countable subadditivity)
Define $B_n = \bigcup_{i=1}^{n} A_i, \forall i \in \mathbb{N}$.

$$\begin{aligned}
B_n \uparrow &\implies \lim_n B_n = \bigcup_{n=1}^{\infty} B_n = \bigcup_{i=1}^{\infty} A_i \\
&\implies \mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_n \mathbb{P}(B_n)
\end{aligned}$$

$$\mathbb{P}(B_n) = \mathbb{P}\left(\bigcup_{i=1}^{n} A_i\right) \le \sum_{i=1}^{n} \mathbb{P}(A_i) \implies \lim_n \mathbb{P}(B_n) \le \sum_{i=1}^{\infty} \mathbb{P}(A_i)$$

□

Remark 1.24. $\Omega = \{\omega_1, ..., \omega_n\}$ is discrete and finite, $\mathcal{S} = 2^\Omega$. It is sufficient to assign $\mathbb{P}$ on $\{\omega_j\}, j = 1, ..., n$. One example is the uniform probability: $\mathbb{P}(\{\omega_j\}) = \frac{1}{n}$.

Example 1.25. (Matching Problem)
Suppose there are m men at a party throwing their hats into a basket. The hats are mixed up and then each man randomly selects a hat. What is the probability that none of the m men selects their own hat?
$\Omega = \{(a_1, ..., a_i, ..., a_m) : a_i \in \{1, ..., m\}, i = 1, ..., m\}$ e.g. $m = 5, (3, 1, 4, 5, 2)$ (men and hats each numbered from 1 to m). This is a permutation, $|\Omega| = m!$. $\mathbb{P}$ is uniform. Define $E_i = \{(a_1, ..., a_m) : a_i = i\}$.
$|E_i| = (m-1)!$ (permutations of $m-1$ elements) $\mathbb{P}(E_i) = \frac{(m-1)!}{m!}$

$$E = \left(\bigcup_{i=1}^{m} E_i\right)^c$$

Inclusion exclusion formula:

$$\mathbb{P}\left(\bigcup_{i=1}^{m} E_i\right) = \sum_{i=1}^{m} \mathbb{P}(E_i) - \sum_{i_1<i_2}^{m} \mathbb{P}(E_{i_1} \cap E_{i_2}) + ... + (-1)^{m+1}\mathbb{P}\left(\bigcap_{i=1}^{m} E_i\right)$$

For $i_1 < ... < i_n, |E_{i_1} \cap ... \cap E_{i_n}| = (m-n)! \implies \frac{(m-n)!}{m!}$.

$$|\{(i_1, ..., i_m) : i_1 < ... < i_n\}| = \binom{m}{n} = \frac{m!}{n!(m-n)!}$$

$$\mathbb{P}\left(\bigcup_{i=1}^{m} E_i\right) = m\frac{(m-1)!}{m!} - \binom{m}{2}\frac{(m-2)!}{m!} + ... + (-1)^{m+1}\frac{1}{m!}$$

$$= \sum_{i=1}^{m}(-1)^{i+1}\binom{m}{i}\frac{(m-i)!}{m!} = \sum_{i=1}^{m}(-1)^{i+1}\frac{1}{i!}$$

$$\mathbb{P}(E) = 1 - \sum_{i=1}^{m}(-1)^{i+1}\frac{1}{i!}$$

What if $m \to \infty$?

$$1 - \lim_{m\to\infty}\left(\sum_{i=1}^{m}(-1)^{i+1}\frac{1}{i!}\right) = 1 + \lim_{m\to\infty}\sum_{i=1}^{m}(-1)^{i}\frac{1}{i!} = \lim_{m\to\infty}\sum_{i=0}^{m}(-1)^{i}\frac{1}{i!}$$

$$= \sum_{i=0}^{\infty}\frac{(-1)^i}{i!} = e^{-1} \approx 0.37$$

Note: combinatorics is Section 1.4 of R-S.

Remark 1.26. For $\Omega = [0,1], \mathbb{R}, \mathbb{R}^n, [a,b] \times [c,d]$, $\mathcal{S} = \mathcal{B}(\Omega)$ we do have a uniform probability.
$\Omega = [0,1]$, uniform $\mathbb{P}$ is such that for every interval $I = [a,b] \in [0,1], \mathbb{P}(I) = |b-a|$. $(\mathbb{P}([0, \frac{1}{2}]) = \frac{1}{2})$.
$\Omega = D \subset \mathbb{R}^n$ bounded, uniform $\mathbb{P}$ is such that $\forall A \in \mathcal{B}(D), \mathbb{P} = \frac{\text{vol}A}{\text{vol}D}$.

Remark 1.27. A countable union of null sets ($A \in \mathcal{S}, \mathbb{P}(A) = 0$ then A is a null set) is again a null set.

Example 1.28. Flip a fair coin infinitely many times. What is the probability that successive heads never occur?

Proof. Consider the first n flips. Define A_n to be the event that no successive heads occur. A_n^c is the event that successive heads occur. Define a_n to be the number of ways of flipping a coin n times but successive heads do not occur.

$$a_n = |A_n|; \quad \mathbb{P}(A_n) = \frac{a_n}{2^n}$$

$$a_1 = 2, a_2 = 3, ...$$

Now suppose that the first flip is a tails. Then for A_n to occur, there must be no successive heads in the other $n-1$ flips. Otherwise heads is flipped first. Then tails must be flipped next and the $n-2$ other flips must not have successive heads. So we have the Fibonacci sequence

$$a_n = a_{n-1} + a_{n-2}$$

note that (a_n) is increasing. Then we have

$$a_n = a_{n-1} + a_{n-2} \leq 2a_{n-1} \implies a_n \leq 3a_{n-2}$$

If $n = 2k$ then we have

$$a_n \leq 3^{k-1}a_2 = 3^k = (\sqrt{3})^n$$

If $n = 2k+1$ then

$$a_n = 3^k a_1 = 2 \cdot 3^k = \frac{2}{\sqrt{3}} \cdot (\sqrt{3})^n$$

$$\implies \frac{a_n}{2^n} \leq \frac{\frac{2}{\sqrt{3}}(\sqrt{3})^k}{2^n} \xrightarrow{n\to\infty} 0$$

$\square$

Example 1.29. Choose two points randomly from $[0,1]$. The two points divide $[0,1]$ into three line segments. What is the probability that the three segments form a triangle? (event A)

Proof.
$$\Omega = \{(x, y) : x, y \in [0, 1]\} = [0, 1] \times [0, 1]; \quad \mathcal{S} = \mathcal{B}(\Omega)$$
If $x < y$ then we must have $x < (y - x) + (1 - y); y - x < x + (1 - y); 1 - y < x + (y - x)$. This gives us the requirement that $x < \frac{1}{2}; y < x + \frac{1}{2}; y > \frac{1}{2}$. Then we have that (x, y) is in the upper left quadrant of $[0, 1]^2$ and also below the line $y = x + \frac{1}{2}$.
Now if $x > y$ we have the same argument in reverse. Then we have two disjoint right triangles with sidelengths $\frac{1}{2}, \frac{1}{2}, \frac{\sqrt{2}}{2}$ so we have $\mathbb{P}(A) = \frac{2 \cdot \frac{1}{2} \cdot \left(\frac{1}{2}\right)^2}{1} = \frac{1}{4}$ □

1.2 Conditional Probability

The study of $(\Omega, \mathcal{S}, \mathbb{P})$ where conditioning on H occurs for some $H \in \mathcal{S}$.

Example 1.30. Toss a fair coin twice.

$$\Omega = \{HH, HT, TH, TT\}$$

$\mathbb{P}$ is the equal probability function.

Now toss a fair coin twice and suppose that H occurs at least once.

$$\Omega = \{HH, HT, TH\}; \quad \mathbb{P}(HH) = \frac{1}{3}$$

Definition 1.31. Let $(\Omega, \mathcal{S}, \mathbb{P})$ be a probability space and let $H \in \mathcal{S}, \mathbb{P}(H) > 0$. Then, $\forall A \in \mathcal{S}$, define
$$\mathbb{P}(A|H) := \frac{\mathbb{P}(A \cap H)}{\mathbb{P}(H)}$$
to be the conditional probability of A conditioning on H. If $\mathbb{P}(H) = 0$ then $\mathbb{P}(\cdot|H)$ is undefined.

Theorem 1.32. *Let $(\Omega, \mathcal{S}, \mathbb{P})$ and H be the same as above. Then the set function $\mathbb{P}(\cdot|H) : A \in \mathcal{S} \mapsto \mathbb{P}(A|H)$ is a probability on $(\Omega, \mathcal{S})$.*

Proof. First, $\forall A \in \mathcal{S}, \mathbb{P}(A|H) \geq 0$ and $\mathbb{P}(\Omega|H) = \frac{\mathbb{P}(\Omega \cap H)}{\mathbb{P}(H)} = 1$. Now take $\{A_i : i \in \mathbb{N}\}$ to be a sequence of disjoint events

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

□

Corollary 1.33. *Let $(\Omega, \mathcal{S}, \mathbb{P})$ and H be the same as above. Define*

$$\mathcal{S}_H = \{A \cap H : A \in \mathcal{S}\}$$

Then $(H, \mathcal{S}_H, \mathbb{P}(\cdot|H))$ is a probability space.

Proof. Exercise. Note: show $\mathcal{S}_H$ is a σ-field, show $\mathbb{P}(\cdot|H)$ is a probability on $(H, \mathcal{S}_H)$. $\square$

Remark 1.34.

$$\mathbb{P}(A|H) = \frac{\mathbb{P}(A \cap H)}{\mathbb{P}(H)} \iff \mathbb{P}(A \cap H) = \mathbb{P}(H)\mathbb{P}(A|H)$$

Theorem 1.35. *(The Multiplication Rule)*
If $A_i \in \mathcal{S}, i = 1, 2, ..., n$ and $\mathbb{P}\left(\bigcap_{i=1}^n A_i\right) > 0$. Then

$$\mathbb{P}\left(\bigcap_{i=1}^n A_i\right) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_2 \cap A_1)... = \prod_{i=1}^n \mathbb{P}\left(A_i | \bigcap_{j=1}^{i-1} A_j\right)$$

Proof.

$$\mathbb{P}\left(\bigcap_{j=1}^{n-1} A_j\right) > 0 \implies \mathbb{P}\left(\bigcap_{j=1}^{k} A_j\right) > 0 \forall k = 1, 2, ..., n-1$$

$$\prod_{i=1}^n \mathbb{P}\left(A_i | \bigcap_{j=1}^{i-1} A_j\right) = \mathbb{P}(A_1)\frac{\mathbb{P}(A_1 \cap A_2)}{\mathbb{P}(A_1)} \cdots \frac{\mathbb{P}\left(\bigcap_{i=1}^n A_i\right)}{\mathbb{P}\left(\bigcap_{i=1}^{n-1} A_i\right)} = \mathbb{P}\left(\bigcap_{i=1}^n A_i\right)$$

$\square$

Example 1.36. Recall the matching problem. Probability of no men receiving their hat is $\sum_{i=0}^m \frac{(-1)^i}{i!}$. What is the probability that exactly k of the m men receive their original hat?

Proof. Fix k spots out of M. Define E to be that event that each man of this set of k men gets his own hat. Define G to be the event that none of the rest of the $(m-k)$ gets his hat back. We then want to find

$$\mathbb{P}(E \cap G) = \mathbb{P}(E)\mathbb{P}(G|E) = \frac{(m-k)!}{m!} \cdot \sum_{i=0}^{m-k} \frac{(-1)^i}{i!}$$

$$\mathbb{P}(\text{exactly } K \text{ of } m \text{ men have matches}) = \binom{m}{k}\frac{(m-k)!}{m!} \cdot \sum_{i=0}^{m-k} \frac{(-1)^i}{i!} = \frac{1}{k!}\sum_{i=0}^{m-k} \frac{(-1)^i}{i!}$$

$\square$

Definition 1.37. Given $(\Omega, \mathcal{S}, \mathbb{P})$ assume that $H_i \in \mathcal{S}, i \in \mathbb{N}$ such that $\forall i \neq j, H_i \cap H_j = \emptyset$ and $\bigcup_{i=1}^{\infty} H_i = \Omega$. This is a partition of Ω.

Proposition 1.38. *Let $\{H_i : i \in \mathbb{N}\}$ be a partition of Ω and assume that $\mathbb{P}(H_i) > 0, \forall i \in \mathbb{N}$. Then $\forall A \in \mathcal{S}$,*

$$\mathbb{P}(A) = \sum_{i=1}^{\infty} \mathbb{P}(A|H_i)\mathbb{P}(H_i)$$

Proof. $A = \bigcup_{i=1}^{\infty}(A \cap H_i)$ and the $(A \cap H_i)$ are disjoint. Thus

$$\mathbb{P}(A) = \sum_{i=1}^{\infty} \mathbb{P}(A \cap H_i) = \sum_{i=1}^{\infty} \mathbb{P}(A|h_i)\mathbb{P}(H_i)$$

$\square$

Theorem 1.39. *(Bayes Rule)*
Let $\{H_i : i \in \mathbb{N}\}$ be a partition with $\mathbb{P}(H_i) > 0, \forall i \in \mathbb{N}$ and $A \in \mathcal{S}$ with $\mathbb{P}(A) > 0$. Then

$$\mathbb{P}(H_i|A) = \frac{\mathbb{P}(H_i)\mathbb{P}(A|H_i)}{\sum_{j=1}^{\infty} \mathbb{P}(H_j)\mathbb{P}(A|H_j)}$$

Remark 1.40. The intuition for this rule is to think of each H_i as a possible cause of A. Then the left hand side of the rule is determining the likelihood of each H_i actually causing A, if A occurs.

Proof.

$$\mathbb{P}(H_i|A) = \frac{\mathbb{P}(H_i \cap A)}{\mathbb{P}(A)} = \frac{\mathbb{P}(H_i)\mathbb{P}(A|H_i)}{\sum_{j=1}^{\infty} \mathbb{P}(H_j)\mathbb{P}(A|H_j)}$$

$\square$

Example 1.41. Three cars, each of which is coloured on both sides. One car has both sides being red, one has both sides black, the third car has one side red and one side black. Choose a car at random, and show one side of the car at random and observe that it is red. What is the probability that the other side of the car chosen is also red?

Proof. Define H_i to be the event that the ith car is chosen $(i = 1, 2, 3)$, A to be the event that the side shown is red. Then we are looking for $\mathbb{P}(H_1|A)$.

$$\mathbb{P}(H_1|A) = \frac{\mathbb{P}(H_1)\mathbb{P}(A|H_1)}{\sum_{j=1}^{3} \mathbb{P}(H_j)\mathbb{P}(A|H_j)} = \frac{\frac{1}{3} \cdot 1}{\frac{1}{3} \cdot 1 + 0 + \frac{1}{3} \cdot \frac{1}{2}} = \frac{2}{3}$$

$\square$

Example 1.42. There are $n+1$ urns marked $0, ..., n$. Each of these urns contains n balls. The kth urn contains k blue balls and $n-k$ yellow balls. An urn is chosen at random, and m draws are made from it, where the ball drawn is always put back before the next draw. Iff all m draws are blue, then what is the probability that the $m+1$ draw is also blue.

Proof. Define H_k the be the event that the kth urn is chosen, $\mathbb{P}(H_k) = \frac{1}{n+1}$, $k = 0, 1, ..., n$. Define A to be the event that the m draws are blue. Define B to be the event that the $m+1$ draw is blue.

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)} = \frac{\sum_{k=0}^{n} \mathbb{P}(H_k)\mathbb{P}(B \cap A|H_k)}{\sum_{j=0}^{n} \mathbb{P}(H_j)\mathbb{P}(A|H_j)} = \frac{\sum_{k=0}^{n} \frac{1}{n+1} \cdot \left(\frac{k}{n}\right)^{m+1}}{\sum_{j=0}^{n} \frac{1}{n+1} \left(\frac{j}{n}\right)^{m}} = \frac{\sum_{k=0}^{n} k^{m+1}}{n \sum_{j=0}^{n} j^m}$$

Now consider what happens as $n \to \infty$. Note the bound

$$\sum_{k=1}^{n} \int_{k-1}^{k} x^{m+1} dx \leq \sum_{k=0}^{n} k^{m+1} \leq \sum_{k=0}^{n} \int_{k}^{k+1} x^{m+1} dx$$

$$\implies \int_{1}^{n} x^{m+1} dx \leq \sum_{k=0}^{n} k^{m+1} \leq \int_{0}^{n+1} x^{m+1} dx$$

$$\implies \frac{n^{m+2}-1}{m+1} \leq \sum_{k=0}^{n} k^{m+1} \leq \frac{(n+1)^{m+2}}{m+2}$$

$$\implies \sum_{k=0}^{n} k^{m+1} = \frac{1}{m+2} n^{m+2} + \mathcal{O}(n^{m+1})$$

and similarly

$$\sum_{j=0}^{n} j^m = \frac{1}{m+1} n^{m+1} + \mathcal{O}(n^m)$$

$$\implies \frac{\sum_{k=0}^{n} k^{m+1}}{n \sum_{j=0}^{n} j^m} = \frac{\sum_{k=0}^{n} k^{m+1} = \frac{1}{m+2} n^{m+2} + \mathcal{O}(n^{m+1})}{\sum_{j=0}^{n} j^m = \frac{1}{m+1} n^{m+1} + \mathcal{O}(n^m)}$$

as $n \to \infty$ this goes to

$$\frac{m+1}{m+2}$$

$\square$

1.3 Independence of Events

Definition *1.43.* Given $(\Omega, \mathcal{S}, \mathbb{P})$; $\{A_1, ..., A_n\} \subset \mathcal{S}$. We say that $\{A_1, ..., A_n\}$ is an independent family (or that they are independent events) if $\forall k = 1, ..., n, \forall 1 \leq i_1 < i_2 < \cdots < i_k \leq n$ we have

$$\mathbb{P}\left(\bigcap_{j=1}^{k} A_{i_j}\right) = \prod_{j=1}^{k} \mathbb{P}(A_{i_j})$$

In particular $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

Definition *1.44.* Given $(\Omega, \mathcal{S}, \mathbb{P})$ let $\mathcal{E} \subset \mathcal{S}$ be a collection of events. $\mathcal{E}$ is an independent family if every finite subcollection of $\mathcal{E}$ is an independent family.

Definition *1.45.* Given $(\Omega, \mathcal{S}, \mathbb{P})$ let $\mathcal{E} \subset \mathcal{S}$ be a collection of events. $\mathcal{E}$ is a pairwise independent family if $\forall A, B \in \mathcal{E}; A \neq B$ we have

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$$

Example *1.46.* Toss a fair coin twice. $\Omega = \{HH, HT, TH, TT\}$. Define the following events.

$$E_1 = \{HH, HT\}; \quad E_2 = \{HH, TH\}; \quad E_3 = \{HH, TT\}$$

Every event has $\mathbb{P}(E_i) = \frac{1}{2} \implies \mathbb{P}(E_i \cap E_j) = \frac{1}{4} = \mathbb{P}(E_i)\mathbb{P}(E_j)$. However, $\mathbb{P}(E_1 \cap E_2 \cap E_3) = \frac{1}{2} \neq \frac{1}{8} = \mathbb{P}(E_1)\mathbb{P}(E_2)\mathbb{P}(E_3)$.

Remark *1.47.* $A, B \in \mathcal{S}$.

- if $A \subset B$ then A and B cannot be independent unless $\mathbb{P}(A) = 0$ or $\mathbb{P}(B) = 1$.
- if $A \cap B = \emptyset$ then A and B are not independent unless $\mathbb{P}(A) = 0$ or $\mathbb{P}(B) = 0$
- if A and B are independent, $\mathbb{P}(B) > 0$ then $\mathbb{P}(A|B) = \mathbb{P}(A)$
- A and B are independent $\iff$ A and B^c are independent $\iff$ A^c and B^c are independent $\iff$ A^c and B are independent

Example *1.48.* An urn contains k blue balls and $n - k$ yellow balls. 2 draws are made from it, where the ball drawn is put back before the next draw.

Proof. Mark the balls $B_1, ..., B_k$ (blue) and $B_{k+1}, ..., B_n$ (yellow).

$$\Omega = \{B_i B_j : 1 \leq i, j \leq n\};$$

$$E_1 = \{B_i B_j : 1 \leq i \leq k, 1 \leq j \leq n\}; E_2 = \{B_i B_j : 1 \leq i \leq n, k+1 \leq j \leq n\}$$

$$E_1 \cap E_2 = \{B_i B_j : 1 \leq i \leq k, k+1 \leq j \leq n\}$$

Now we have

$$\mathbb{P}(E_1) = \frac{|E_1|}{|\Omega|} = \frac{kn}{n^2} = \frac{k}{n}$$

$$\mathbb{P}(E_2) = \frac{|E_2|}{|\Omega|} = \frac{(n-k)n}{n^2} = \frac{n-k}{n}$$

$$\mathbb{P}(E_1 \cap E_2) = \frac{|E_1 \cap E_2|}{|\Omega|} = \frac{(n-k)k}{n^2} = \mathbb{P}(E_1)\mathbb{P}(E_2)$$

Thus E_1, E_2 are independent □

Example 1.49. An urn contains k blue balls and $n-k$ yellow balls. 2 draws are made from it, where the ball drawn is removed before the next draw.

Proof. Mark the balls $B_1, ..., B_k$ (blue) and $B_{k+1}, ..., B_n$ (yellow).

$$\Omega = \{B_iB_j : 1 \leq i, j \leq n; j \neq i\};$$

$$E_1 = \{B_iB_j : 1 \leq i \leq k, 1 \leq j \leq n; j \neq i\}; E_2 = \{B_iB_j : 1 \leq i \leq n, k+1 \leq j \leq n; j \neq i\}$$

$$E_1 \cap E_2 = \{B_iB_j : 1 \leq i \leq k, k+1 \leq j \leq n\}$$

Now we have

$$\mathbb{P}(E_1) = \frac{|E_1|}{|\Omega|} = \frac{k(n-1)}{n(n-1)} = \frac{k}{n}$$

$$\mathbb{P}(E_2) = \frac{|E_2|}{|\Omega|} = \frac{(n-k)k + (n-k)(n-k-1)}{n(n-1)} = \frac{n-k}{n}$$

$$\mathbb{P}(E_1 \cap E_2) = \frac{|E_1 \cap E_2|}{|\Omega|} = \frac{(n-k)k}{n(n-1)} > \mathbb{P}(E_1)\mathbb{P}(E_2)$$

So now E_1, E_2 are not independent. □

2 Discrete and Continuous Random Variables

2.1 Distributions

Definition 2.1. If Ω is finite – i.e. $\Omega - \{\omega_1, ..., \omega_n\}$ – or countable – $\Omega = \{\omega_i : i \in \mathbb{N}\}$ – and $\mathcal{S} = 2^\Omega$ then it suffices to assign on $\{\omega\}, \omega \in \Omega$ then $\forall A \in \mathcal{S}, \mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\{\omega\})$. In this case, $(\Omega, \mathcal{S}, \mathbb{P})$ is a discrete probability space, $\mathbb{P}$ a discrete probability or a discrete distribution.

Definition 2.2. The function $f : \Omega \to [0,1]$ given by $f(\omega) = \mathbb{P}(\{\omega\}), \forall \omega \in \Omega$ is the probability mass function (PMF) of $\mathbb{P}$.

Proposition 2.3. *Given discrete Ω and $\mathcal{S} = 2^\Omega$ we have $f : \Omega \to [0,1]$ is the PMF of some $\mathbb{P}$ on $(\Omega, \mathcal{S})$ if and only if $\sum_{\omega \in \Omega} f(\omega) = 1$.*

Proof. The forward direction is obvious, the reverse comes from defining $\mathbb{P}(\{\omega\}) = f(\omega)$. $\square$

The following are examples of discrete distributions – $\mathcal{S} = 2^\Omega$

Example 2.4. $\Omega = \{\omega_1, ..., \omega_n\}$ then $\mathbb{P}$ is the discrete uniform distribution if the PMF is $f(\omega_i) = \frac{1}{n}, i = 1, ..., n$

Example 2.5. $\omega_0 \in \Omega$. $\mathbb{P}$ is the Dirac distribution at ω_0 if PMF is $f(\omega_0) = 1, f \equiv 0$ on $\Omega - \{\omega_0\}$

Example 2.6. $\Omega = \{0, 1\}; p \in (0, 1)$. $\mathbb{P}$ is the Bernoulli distribution denoted by $B(p)$ if the PMF is $f(1) = p, f(0) = 1 - p$.
$B(p)$ is used to model trials with only two outcomes. (success/failure)

Example 2.7. $\Omega = \{0, 1, ..., n\}, p \in (0, 1)$. $\mathbb{P}$ is the binomial distribution denoted by $B(n, p)$ if the PMF is $F(k) = \binom{n}{k} p^k (1-p)^{n-k}; k = 0, 1, ..., n$. We have $\sum_{k=0}^n f(k) = \sum_{k=0}^n \binom{n}{k} p^k (1-p)^{n-k} = (p + 1 - p)^n = 1$ so f is a PMF.
$B(n, p)$ models the number of successes in n independent Bernoulli trials. $\mathbb{P}$(success in k out of n trials) $= \binom{n}{k} p^k (1-p)^{n-k} \to B(n, p)$

Example 2.8. $\Omega = \{0\} \cup \mathbb{N}; \lambda > 0$. $\mathbb{P}$ is the Poisson distribution denoted by $P(\lambda)$. The PMF is $f(k) = e^{-\lambda} \frac{\lambda^k}{k!}, k = 0, 1, 2, ...$ We have

$$\sum_{k=0}^{\infty} f(k) = e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1$$

models the number of occurrences of rare events.

Example 2.9. Study the number of Pokémon you find in a day. Assume that the likelihood of you finding a Pokémon in a day is $\lambda > 0$. Equally divide the day into n time intervals with n large. Then within each time interval your chance of finding one Pokémon is $\frac{\lambda}{n}$.
The number of Pokémon found in a day is the number of successes you have in a day $\approx B\left(n, \frac{\lambda}{n}\right)$.
PMF of $B\left(n, \frac{\lambda}{n}\right)$:

$$f(k) = \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} = \frac{n!}{k!(n-k)!} \cdot \frac{\lambda^k}{n^k} \cdot \frac{(n-\lambda)^{n-k}}{n^{n-k}}$$

$$= \frac{\lambda^k}{k!} \cdot \frac{n(n-1)\cdots(n-k+1)(n-\lambda)^n}{n^n (n-\lambda)^k}$$

$$= \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^n \cdot \left(\frac{n}{n-\lambda}\right) \cdot \left(\frac{n-1}{n-\lambda}\right) \cdots \left(\frac{n-k+1}{n-\lambda}\right) \xrightarrow{n \to \infty} \frac{\lambda^k}{k!} e^{-\lambda} = P(\lambda)$$

So $P(\lambda)$ is the limit of $B\left(n, \frac{\lambda}{n}\right)$.

Definition 2.10. Consider the probability $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{Q})$ – i.e. $\mathbb{Q}$ is a probability on the space.
The distribution function (DF) F of $\mathbb{Q}$ is defined as $F : \mathbb{R} \to [0,1]$ and $F(x) = \mathbb{Q}((-\infty, x]), \forall x \in \mathbb{R}$.
Properties of F:

- F is nondecreasing on $\mathbb{R}$
- $\lim_{x\to-\infty} F(x) = 0; \lim_{x\to\infty} F(x) = 1$ – this is because $\{(-\infty, -n] : n \in \mathbb{N}\}$ is a monotonic (non increasing) sequence, and $\bigcap_{n=1}^{\infty}(-\infty, -n] = \emptyset$. Similar for the other limit.
- F is continuous from the right i.e., $\forall a \in \mathbb{R}, \lim_{x\to a^+} F(x) = F(a)$

Remark 2.11. Fix $a \in \mathbb{R}$, and let $\{x_n\}$ such that $x_n \to a^+$ as $n \to \infty$.

$$(-\infty, x_n] \downarrow \bigcap_{n=1}^{\infty}(-\infty, x_n] = (-\infty, a]$$

$$\implies \lim_{n\to\infty} \mathbb{Q}((-\infty, x_n]) = \mathbb{Q}((-\infty, a]) \implies \lim_{n\to\infty} F(x_n) = F(a)$$

Remark 2.12. F is not necessarily left continuous.

Proposition 2.13. *If $F : \mathbb{R} \to [0,1]$ satisfies (i),(ii),(iii) from the above then F is the DF for some probability $\mathbb{Q}$ on $(\mathbb{R}, \mathcal{B}(\mathbb{R})$.*

Proof. Proof is omitted here. The idea is take F that satisfies the three properties and define $\mathbb{Q} : (a,b] \mapsto F(b) - F(a)$. Argue that $\mathbb{Q}$ is a probability on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Check that $\mathbb{Q} \geq 0, \mathbb{Q}(\mathbb{R}) = 1, \mathbb{Q}$ is countably additive. □

Remark 2.14. Many discrete distributions can be thought of as distributions on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Then the DF of a discrete distribution is a jump function/step function/piece-wise continuous function.

Example 2.15. Binomial distribution $B(n,p), \Omega\{0,1,...,n\} \subset \mathbb{R}, \mathcal{S} = 2^{\Omega} \subset \mathcal{B}(\mathbb{R})$. The DF of $B(n,p)$:

$$F(x) = \mathbb{P}((-\infty, x]) = \sum_{k\in\{0,1,\ldots,n\}, k\leq x} f(k) = \sum_{k\in\{0,\ldots,n\}, k\leq x} \binom{n}{k} p^k (1-p)^{n-k}$$

where f is the PMF of $B(n,p)$.

- if $x < 0$ then $F(x) = 0$.
- if $0 \leq x < 1$ then $F(x) = (1-p)^n$

- if $1 \leq x < 2$ then $F(x) = (1-p)^n + np(1-p)^{n-1}$

Example 2.16. Poisson distribution $P(\lambda), \Omega\{0, 1, 2, ...\} \subset \mathbb{R}, \mathcal{S} = 2^\Omega \subset \mathcal{B}(\mathbb{R})$. The DF of $B(n, p)$:

$$F(x) = \mathbb{P}((-\infty, x]) = \sum_{k \in \mathbb{N}, k \leq x} f(k) = \sum_{k \in \mathbb{N}, k \leq x} e^{-\lambda} \frac{\lambda^k}{k!}$$

where f is the PMF of $P(\lambda)$.

- if $x < 0$ then $F(x) = 0$.
- if $0 \leq x < 1$ then $F(x) = e^{-\lambda}$
- if $1 \leq x < 2$ then $F(x) = e^{-\lambda} + \lambda e^{-\lambda}$
- $\forall x \in \mathbb{R}, F(x) < 1$ because there will always be values greater than it, and the jumps will get smaller and smaller

Remark 2.17. PMF is more useful for discrete distributions.

Example 2.18. Define:

$$F(x) = \begin{cases} 0: & \text{if } x \leq 0 \\ x: & \text{if } x \in [0, 1] \\ 1: & \text{if } x \geq 1 \end{cases} \tag{2.1}$$

F is a DF for some $\mathbb{Q}$ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

- $\mathbb{Q}((-\infty, x]) = F(x)$. Check that $\mathbb{Q}(B) = 0$ if $B \subset (-\infty, 0]$.
- $\forall$ intervals $I \subset [0, 1]$ then $\mathbb{Q}(I) =$ length of I
- $\forall A \subset \mathbb{R}$ such that $[0, 1] \subset A, \mathbb{Q}(A) = 1$

Q is the uniform distribution on $[0, 1] : ([0, 1], \mathcal{B}([0, 1]), \mathbb{Q})$

Definition 2.19. Given $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{Q})$ and the DF of $\mathbb{Q}$, say F, $\mathbb{Q}$ is a continuous distribution if F is absolutely continuous, i.e., $\exists f : \mathbb{R} \to [0, \infty)$ such that $\forall x \in \mathbb{R}$

$$F(x) = \int_{-\infty}^{x} f(t) dt$$

Then f is the probability density function of $\mathbb{Q}$.

Remark 2.20. If $\mathbb{Q}$ is a continuous distribution with DF $F(x)$ and PDF $f(t)$

- $\int_{-\infty}^{\infty} f(t) dt = \lim_{x \to \infty} F(x) = 1$

- $\forall a \leq b, \mathbb{Q}((a,b]) = F(b) - F(a) = \int_a^b f(t)dt$
- F is continuous everywhere, i.e. $\forall a \in \mathbb{R}, \lim_{x \to a} F(x) = F(a)$. Also F is differentiable almost everywhere, i.e. $F'(x)$ exists for almost every $x \in \mathbb{R}$ and $F'(x) = f(x)$ for almost every x as well. However, if f is continuous at x then $F'(x) = f(x)$

Example 2.21. Consider again the DF

$$F(x) = \begin{cases} 0: & \text{if } x \leq 0 \\ x: & \text{if } x \in [0,1] \\ 1: & \text{if } x \geq 1 \end{cases} \tag{2.2}$$

Then we have the PDF

$$f(x) = \begin{cases} 0: & \text{if } x < 0 \\ 1: & \text{if } x \in (0,1) \\ 0: & \text{if } x > 1 \end{cases} \tag{2.3}$$

Example 2.22. Normal (Gaussian) distribution denoted by $N(0,1)$ is a continuous distribution. $N(0,1)$ has DF, PDF

$$F(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt; \quad f(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$$

Remark 2.23. PDF does not exist for discrete distributions.

Remark 2.24. There are distributions in between the discrete and continuous types, e.g.

$$F(x) = \begin{cases} 0: & \text{if } x < 0 \\ \frac{1}{4} + \frac{1}{2}x: & \text{if } 0 \geq x < \frac{1}{2} \\ \frac{3}{4}: & \text{if } \frac{1}{2} \geq x < 1 \\ 1 - \frac{1}{8}e^{1-x}: & \text{if } x \geq 1 \end{cases} \tag{2.4}$$

Remark 2.25. Exercise: Find $a \in (0,1)$, continuous F_1, discrete F_2 such that $F(x) = aF_1(x) + (1-a)F_2(x)$.

Remark 2.26. There are DF $F(x)$ such that $F(x)$ is continuous but $F(x)$ is not absolutely continuous.

Example 2.27. Cantor function.

Remark 2.28. If $\mathbb{Q}$ is a continuous distribution then $\forall x \in \mathbb{R}, \mathbb{Q}(\{x\}) = 0$.

Remark 2.29. View $f(x)$ as the probability near point x.

$$\mathbb{Q}(x - \epsilon, x + \epsilon)) = \int_{x-\epsilon}^{x+\epsilon} f(t)dt \approx f(x)2\epsilon$$

2.2 Random Variables

2.2.1 Discrete Random Variables

Definition 2.30. Let Ω be a sample space, $\mathcal{S}$ a σ-field on Ω. A function $X : \Omega \to \mathbb{R}$ is called a random variable (RV) if

$$\forall B \in \mathcal{B}(\mathbb{R}), X^{-1}(B) := \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{S}$$

Remark 2.31. Properties of a random variable:

- if Ω is discrete and $\mathcal{S} = 2^{\Omega}$ then any function $X : \Omega \to \mathbb{R}$ is a RV
- if $\Omega = \mathbb{R}, \mathcal{S} = \mathcal{B}(\mathbb{R})$ then $\exists$ a function $X : \mathbb{R} \to \mathbb{R}$ that is not a RV, e.g., for $A \subset \mathbb{R}$ define $X = \chi_A$. Then we have X is a RV $\iff A \in \mathcal{B}(\mathbb{R})$. Note, this definition is:

$$X(x) = \chi_A(x) = \begin{cases} 1 : & \text{if } x \in A \\ 0 : & \text{otherwise} \end{cases} \tag{2.5}$$

Now $\forall B \in \mathcal{B}(\mathbb{R})$

$$X^{-1}(B) = \begin{cases} \emptyset : & \text{if } 0,1 \notin B \\ A^c : & \text{if } 0 \in B, 1 \notin B \\ A : & \text{if } 1 \in B, 0 \notin B \\ \mathbb{R} : & \text{if } 0,1 \in B \end{cases} \tag{2.6}$$

Example 2.32. Examples of a RV

- roll a pair of dice. $X =$ sum of the face values is a RV
- randomly choose two points a, b from $[0,1]$, $X = |a - b|$
- looking for Pokémon in a day, $X =$ the time that you find the fifth Pokémon

Remark 2.33. Exercise: Verify that for every function $X : \Omega \to \mathbb{R}, A, B \subset \mathbb{R}, \{B_i : i \in \mathbb{N}\} \subset \mathbb{R}$

- $X^{-1}(B^c) = (X^{-1}(B))^c$
- $X^{-1}(A \cap B) = (X^{-1}(A)) \cap (X^{-1}(B))$; $X^{-1}(A \cup B) = (X^{-1}(A)) \cup (X^{-1}(B))$
- $X^{-1}\left(\bigcap_{i=1}^{\infty} B_i\right) = \bigcap_{i=1}^{\infty}(X^{-1}(B_i))$; $X^{-1}\left(\bigcup_{i=1}^{\infty} B_i\right) = \bigcup_{i=1}^{\infty}(X^{-1}(B_i))$

Theorem 2.34. *Given $(\Omega, \mathcal{S})$ and $X : \Omega \to \mathbb{R}$, X is a RV if and only if $\forall x \in \mathbb{R}$*

$$\{X \leq x\} = \{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{S}$$

Proof. ($\Longrightarrow$) by definition of a RV.
($\Longleftarrow$) Define $\mathcal{D} = \{B \subset \mathbb{R} : X^{-1}(B) \in \mathcal{S}\}$. Show that $\mathcal{D}$ is a σ-field.

- $X^{-1}(\mathbb{R}) = \Omega \in \mathcal{S}$
- if $B \in \mathcal{D}$ then $X^{-1}(B) \in \mathcal{S} \implies X^{-1}(B^c) = (X^{-1}(B))^c \in \mathcal{S} \implies B^c \in \mathcal{D}$
- if $B_i \in \mathcal{D}, i \in \mathbb{N}$ then $\bigcup_{i=1}^{\infty} B_i \in \mathcal{D}$

Now we have that $\mathcal{D}$ is a σ-field.
We know that

$$\forall x \in \mathbb{R}, (-\infty, x] \in \mathcal{D} \implies \{(-\infty, x] : x \in \mathbb{R}\} \subset \mathcal{D}$$

It suffices to show that $\sigma(\{(-\infty, x] : x \in \mathbb{R}\}) = \mathcal{B}(\mathbb{R}) = \sigma(\{(a, b) : a < b; a, b \in \mathbb{R}\})$.
Since $(-\infty, x] \in \mathcal{B}(\mathbb{R})$, LHS $\subset$ RHS.
$\forall a, b \in \mathbb{R}, (a, b] = (-\infty, b] - (-\infty, a]$ and $(a, b) = \bigcup_{n=1}^{\infty} \left(a, b - \frac{1}{n}\right)$ so we have equality. $\square$

Corollary 2.35. *IF X is a RV on $(\Omega, \mathcal{S})$ then*

$$aX+b \; (a, b \in \mathbb{R}, a \neq 0); \quad |X|, \quad X^2; \quad \frac{1}{X} \; (X \neq 0); \quad \sqrt{X} \; (X \geq 0); \quad e^X; \quad \log X \; (X > 0)$$

are all RV.

Proof. First, $aX + b$. $\forall x \in \mathbb{R}$,

$$\{aX + b \leq x\} = \{\omega \in \Omega : aX + b \leq x\}$$

$$= \{\omega \in \Omega : aX \leq x - b\}$$

WLOG $a < 0$

$$= \{\omega \in \Omega : X \geq \frac{x-b}{a}\} = \{X \geq \frac{x-b}{a}\} \in \mathcal{S}$$

Now, $|X|$. $\forall x \in [0, \infty)$

$$\{|X| \leq x\} = \begin{cases} \{-x \leq X \leq x\} = X^{-1}([-x, x]) \in \mathcal{S} : & \text{if } x \geq 0 \\ \{|X| \leq x\} = \emptyset \in \mathcal{S} : & \text{otherwise} \end{cases} \tag{2.7}$$

Others are similar. $\square$

Proposition 2.36. *Given $(\Omega, \mathcal{S})$ and a RV $X : \Omega \to \mathbb{R}$, let $g : \mathbb{R} \to \mathbb{R}$ – a RV – such that $\forall B \in \mathcal{B}(\mathbb{R}), g^{-1}(B) \in \mathcal{B}(\mathbb{R})$. Then $g \circ X : \Omega \to \mathbb{R}$ is a RV.*

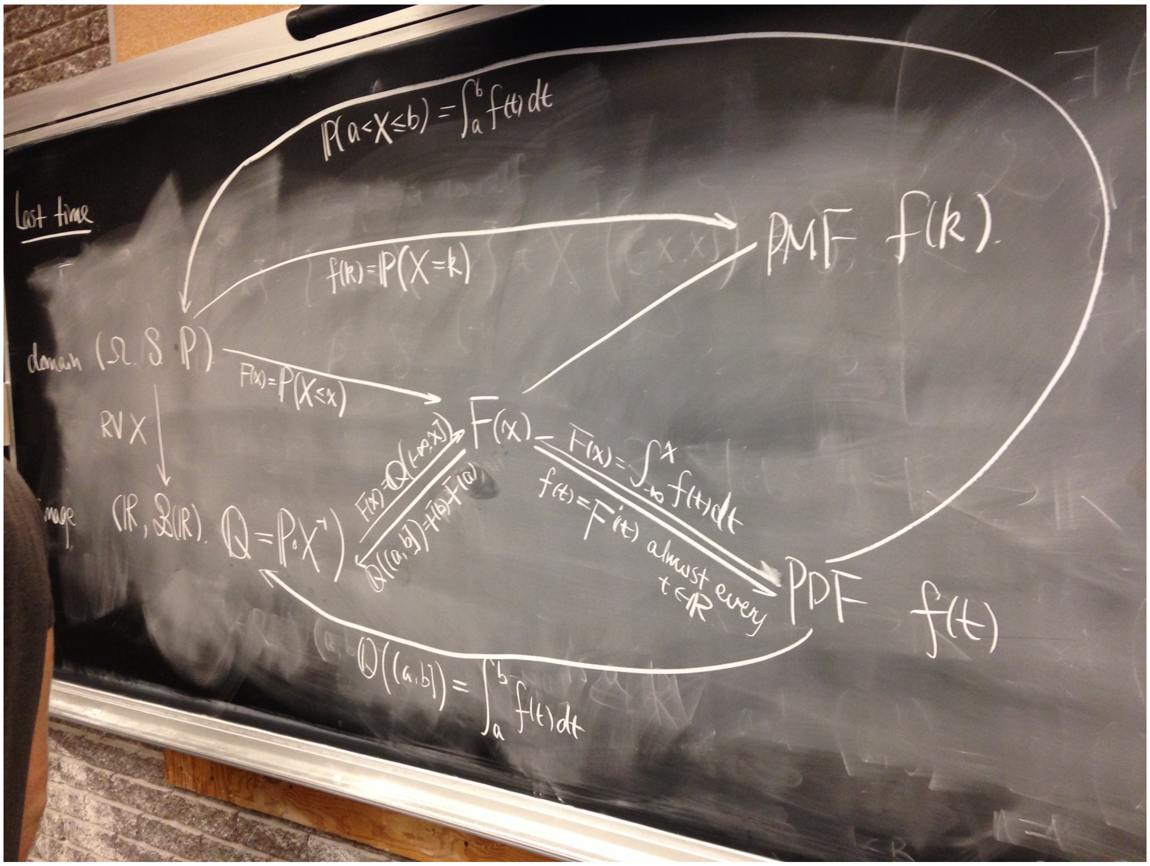
Proof. Exercise. $\square$

Remark 2.37. $\exists A \subset \mathbb{R}$ such that $A \notin \mathcal{B}(\mathbb{R})$. If $g = \chi_A$ then $g \circ X$ is not a RV.

Definition 2.38. Given $(\Omega, \mathcal{S}, \mathbb{P})$ let $X : \Omega \to \mathbb{R}$ be a RV. Then define the following set function on $\mathcal{B}(\mathbb{R})$

$$\forall B \in \mathcal{B}(\mathbb{R}), \mathbb{Q}(B) := \mathbb{P}(X^{-1}(B))$$

$\mathbb{Q}$ is the distribution of RV X. If F is the DF of $\mathbb{Q}$, i.e. $F(x) = \mathbb{Q}((-\infty, x]), \forall x \in \mathbb{R}$. Then F is also called the DF of X. $F(x) = \mathbb{P}(X \leq x)$.



Theorem 2.39. *If $\mathbb{Q}$ is a probability distribution on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ $\iff$ $\exists(\Omega, \mathcal{S}, \mathbb{P})$ and a RV $X : \Omega \to \mathbb{R}$ such that $\mathbb{Q} = \mathbb{P} \circ X^{-1}$.*

Proof. ($\Longleftarrow$) Check the definition.

- $\forall B \in \mathcal{B}(\mathbb{R}), \mathbb{Q}(B) = \mathbb{P}(X^{-1}(B)) \geq 0$
- $\mathbb{Q}(\mathbb{R}) = \mathbb{P}(X^{-1}(\mathbb{R})) = \mathbb{P}(\Omega) = 1$
- $\forall B_i \in \mathcal{B}(\mathbb{R}), i \in \mathbb{N}$ with the B_i's disjoint then

$$\mathbb{Q}\left(\bigcup_{i=1}^{\infty} B_i\right) = \mathbb{P}\left(X^{-1}\left(\bigcup_{i=1}^{\infty} B_i\right)\right) = \mathbb{P}\left(\bigcup_{i=1}^{\infty} X^{-1}(B_i)\right)$$

$$= \sum_{i=1}^{\infty} \mathbb{P}(X^{-1}(B_i)) = \sum_{i=1}^{\infty} \mathbb{Q}(B_i)$$

($\implies$) Exercise.
Take $(\Omega, \mathcal{S}, \mathbb{P}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{Q})$ Define $X = Id, X(x) = x, \forall x \in \mathbb{R}$.
Reference: Skorohod's Representation $\square$

Definition 2.40. We say that X is a discrete uniform/Dirac/Bernoulli/Binomial/Poisson/.../uniform on [0,1]/normal... RV if the distribution $\mathbb{Q}$ of X discrete uniform/Dirac/Bernoulli/Binomial/Poisson/.../uniform on [0,1]/normal...

Definition 2.41. If the distribution $\mathbb{Q}$ of X is a continuous(discrete) distribution with PDF $f(x)$ (PMF $f(k)$) then f is also called the PDF(PMF) of X.

Definition 2.42. A transformation of a RV is the composition of a RV, X, with a Borel function, g, i.e., $g \circ X$. The distribution $\mathbb{Q}_{g\circ X}$ of $g \circ X$ is given by

$$B \in \mathcal{B}(\mathbb{R}) \mapsto \mathbb{Q}_{g\circ X}(B) = \mathbb{P}((g \circ X)^{-1}(B)) = \mathbb{P}(X \in g^{-1}(B)) = \mathbb{Q}_X(g^{-1}(B))$$

Remark 2.43. If X is a discrete RV, i.e., $X \in \{x_n : n \in \mathbb{N}\}$ (X takes values in $\{x_n\}$) then $\forall B \in \mathcal{B}(\mathbb{R})$

$$\mathbb{Q}_{g\circ X}(B) = \mathbb{P}(X \in g^{-1}(B)) = \sum_{\{n\in\mathbb{N}:x_n\in g^{-1}(B)\}} \mathbb{P}(X = x_n) = \sum_{\{n\in\mathbb{N}:x_n\in g^{-1}(B)\}} f_X(x_n)$$

f_X is the PMF of X.

Remark 2.44. Note that if X is discrete, then $g \circ X$ is also a discrete RV and $g \circ X \in \{g(x_n) : n \in \mathbb{N}\}$. If $f_{g\circ X}$ is the PMF of $g \circ X$ then

$$f_{g\circ X}(g(x_n)) = \mathbb{Q}_{g\circ X}(\{g(x_n)\}) = \sum_{\{k\in\mathbb{N}:g(x_k)=g(x_n)\}} f_X(x_k)$$

Example 2.45. X is a $P(\lambda)$ RV. $Y = e^X$ is a discrete RV. $X \in \{0, 1, 2, 3, ...\}; Y \in \{1, e, e^1, e^2, e^3, ...\}$.

$$f_Y(e^k) = \sum_{\{n\in\mathbb{N}:e^n=e^k\}} f_X(n) = f_X(k) = e^{-\lambda}\frac{\lambda^k}{k!}$$

Example 2.46. X is a $B(4, p)$ RV. $X \in \{0, 1, 2, 3, 4\}; Y = (X - 2)^2 \in \{4, 1, 0\}$. $g(x) = (x - 2)^2$

$$f_Y(4) = f_X(0) + f_X(4) = \binom{4}{0}p^0(1-p)^4 + \binom{4}{4}p^4(1-p)^0$$

$$f_Y(1) = f_X(1) + f_X(3) = \binom{4}{1}p^1(1-p)^3 + \binom{4}{3}p^3(1-p)^1$$

$$f_Y(0) = f_X(2) = \binom{4}{2}p^2(1-p)^2$$

2.2.2 Expectation of a Discrete Random Variable

Definition 2.47. Assume that X is a discrete RV on $(\Omega, \mathcal{S}, \mathbb{P})$ and $X \in \{x_n : n \in \mathbb{N}\}$ with PMF of X being $f_X(x_n) = \mathbb{P}(X = x_n), \forall n \in \mathbb{N}$. If

$$\sum_{n=1}^{\infty} |x_n| f_X(x_n) = \sum_{n=1}^{\infty} |x_n| \mathbb{P}(X = x_n) < \infty$$

then we say that the expectation of X exists and it is defined to be

$$\mathbb{E}[X] = \sum_{n=1}^{\infty} x_n f_X(x_n) = \sum_{n=1}^{\infty} x_n \mathbb{P}(X = x_n)$$

If $\sum_{n=1}^{\infty} |x_n| \mathbb{P}(X = x_n) = \infty$ then we say that the expectation does not exist.

Remark 2.48. X is a RV with PMF

$$f_X\left((-1)^{j+1}\frac{3^j}{j}\right) = \frac{2}{3^j}, j = 1, 2, 3, ...$$

Check that it is a PMF, i.e., $\sum_{j=1}^{\infty} \frac{2}{3^j} = 1$. Now,

$$\sum_{j=1}^{\infty} x_j \frac{2}{3^j} = 2\sum_{j=1}^{\infty} (-1)^{j+1} \frac{3^j}{j}\frac{2}{3^j} = 2\sum_{j=1}^{\infty} (-1)^{j+1}\frac{1}{j} = 2\ln 2$$

Note $\ln(1 + x) = \sum_{n=1}^{\infty} (-1)^{n+1} \frac{x^n}{n}$ when $0 < x \leq 1$. However:

$$\sum_{j=1}^{\infty} \frac{2}{j} = \infty$$

by comparison with the integral $\int_1^{\infty} \frac{1}{x} dx$. Thus, $\mathbb{E}[X]$ does not exist.

Remark 2.49.

- $\mathbb{E}[X]$ exists if and only if $\mathbb{E}[|X|] < \infty$
- If $X \geq 0$ then if $\mathbb{E}[X]$ exists, $\mathbb{E}[X] = \mathbb{E}[|X|]$
- If $\mathbb{E}[X]$ exists then $|\mathbb{E}[X]| \leq \mathbb{E}[|X|]$
- If X is a finite RV then $\mathbb{E}[X], \mathbb{E}[|X|]$ always exists

The expectation is the "average" of X or "mean" of X.

Example 2.50. If X is the discrete uniform RV on $\{x_1, ..., x_n\}$ then

$$\mathbb{E}[X] = \sum_{i=1}^{n} x_i \mathbb{P}(X = x_i) = \frac{\sum_{i=1}^{n} x_i}{n}$$

This is the average.

Example 2.51. If X is the Bernoulli RV ($p \in (0,1)$), $X \in \{0,1\}$ then

$$\mathbb{E}[X] = 0(1-p) + 1 \cdot p = p$$

Example 2.52. If X is $B(n,p)$, $X \in \{0,...,n\}$ then

$$\mathbb{E}[X] = \sum_{k=0}^{n} k\mathbb{P}(X=k) = \sum_{k=0}^{n} k\binom{n}{k}p^k(1-p)^{n-k}$$

$$= np\sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!(n-1-(k-1))!}p^{k-1}(1-p)^{n-1-(k-1)}$$

Now setting $l = k-1$

$$= np\sum_{l=0}^{n-1}\binom{n-1}{l}p^l(1-p)^{n-1-l} = np$$

Since we have that $\binom{n-1}{l}p^l(1-p)^{n-1-l}$ is the PMF of $B(n-1,p)$.

Example 2.53. If X is $P(\lambda)$, $X \in \{0,1,2,...\}$, $X \geq 0$ then

$$\mathbb{E}[|X|] = \mathbb{E}[X] = \sum_{k=0}^{\infty} k\mathbb{P}(X=k) = \sum_{k=0}^{\infty} k\frac{\lambda^k}{k!}e^{-\lambda}$$

$$\lambda\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!}e^{-\lambda}$$

$$\lambda\sum_{l=0}^{\infty}\frac{\lambda^{l}}{(l)!}e^{-\lambda} = \lambda$$

Since $\frac{\lambda^l}{(l)!}e^{-\lambda}$ is the PMF of $P(\lambda)$.

Remark 2.54. $A \in \mathcal{S}, X = \chi_A, X \in \{0,1\}$. Then

$$\mathbb{E}[X] = 1\mathbb{P}(X=1) = \mathbb{P}(A)$$

Theorem 2.55. *$X : \Omega \to \mathbb{R}$ is a discrete RV, $X \in \{x_n : n \in \mathbb{N}\}, g : \mathbb{R} \to \mathbb{R}$ Borel, $Y = g \circ X$ is a discrete RV. Then*

$$\mathbb{E}[g \circ X] exists \iff \sum_{n=1}^{\infty} |g(x_n)|\mathbb{P}(X = x_n) < \infty$$

and in this case

$$\mathbb{E}[g \circ X] = \mathbb{E}[Y] = \sum_{n=1}^{\infty} g(x_n)\mathbb{P}(X = x_n)$$

Proof. $Y \in \{y_k : k \in \mathbb{N}\} = \{g(x_n) : n \in \mathbb{N}\}$.

$$\mathbb{E}[|Y|] = \sum_{k=1}^{\infty} |y_k| \mathbb{P}(Y = y_k) = \sum_{k=1}^{\infty} |y_k| \sum_{\{n:g(x_n)=y_k\}} \mathbb{P}(X = x_n)$$

$$= \sum_{k=1}^{\infty} \sum_{\{n:g(x_n)=y_k\}} |g(x_n)| \mathbb{P}(X = x_n) = \sum_{n=1}^{\infty} |g(x_n)| \mathbb{P}(X = x_n)$$

since $\mathbb{P}(Y = y_k) = \mathbb{P}(g \circ X = y_k) = \mathbb{P}(X \in g^{-1}(\{y_k\}))$ □

Corollary 2.56. *If $\mathbb{E}[X]$ exists then $\forall a, b, \in \mathbb{R}$, $\mathbb{E}[aX + b]$ exists and*

$$a\mathbb{E}[X] + b$$

In particular, if $\mathbb{E}[X] = m$

$$\mathbb{E}[X - m] = 0$$

Proof. $g(x) = ax + b$, $X \in \{x_n : n \in \mathbb{N}\}$.

$$\sum_n |ax + b| \mathbb{P}(X = x_n) \leq \sum_n (|a| \cdot |x_n| + |b|) \mathbb{P}(X = x_n) = |a| \sum_n |x_n| \mathbb{P}(X = x_n) + |b| < \infty$$

So $\mathbb{E}[aX + b]$ exists.

$$\mathbb{E}[aX + b] = \sum_n (ax_n + b) \mathbb{P}(X = x_n) = a\mathbb{E}[X] + b$$

□

Example 2.57. Let X be a $B(n, p)$ RV. Compute $\mathbb{E}\left[\frac{1}{x+1}\right]$. $g(x) = \frac{1}{x+1}$.

$$\mathbb{E}\left[\frac{1}{x+1}\right] = \sum_{k=0}^{n} \frac{1}{k+1} \binom{n}{k} p^k (1-p)^{n-k}$$

Exercise: show this is equal to

$$\frac{1 - (1-p)^{n+1}}{(n+1)p}$$

Note that $\frac{1}{k+1}\binom{n}{k} = \frac{n!}{(k+1)!(n-k)!}$

Definition 2.58. Let X be a RV for $k \in \{0, 1, 2, ...\}$. If $\mathbb{E}[X^k]$ exists then we can find $\mathbb{E}[X^k]$ is called the kth moment of X, M_k.

$$\mathbb{E}[X^0] = 1$$

$$\mathbb{E}[X^1] = \mathbb{E}[X]$$

Remark 2.59. X is a discrete RV. $X \in \{x_n : n \in \mathbb{N}\}$. If $\sum_n |x_n^k| \mathbb{P}(X = x_n) < \infty$ then $M_k = \sum_n x_n^k \mathbb{P}(X = x_n)$.

Remark 2.60. If I know all the moments of X then I know completely the distribution of X.

Example 2.61. If $X \in \{x_1, ..., x_n\}$ denote $p_n = \mathbb{P}(X = x_n)$. I know M_k for all the ks so we have the system of linear equations

$$M_1 = \mathbb{E}[X] = \sum_{j=1}^{n} x_j p_j$$

$$M_2 = \mathbb{E}[X^2] = \sum_{j=1}^{n} x_j^2 p_j$$

$$...$$

$$M_n = \mathbb{E}[X^n] = \sum_{j=1}^{n} x_j^n p_j$$

which can be solved for each p_i.

Definition 2.62. X is a RV with $\mathbb{E}[X^2] < \infty$. The variance of X denoted by $Var(X)$ is defined by

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

Remark 2.63. $Var(X) \geq 0$. Variance measures "how far/much" the distribution of X from $\mathbb{E}[X]$.

Remark 2.64. The variance of X is 0 if and only if X is a Dirac RV.

Example 2.65. $X \in \{x_1, ..., x_n\}$ a discrete uniform RV.

$$M_2 = \mathbb{E}[X^2] = \sum_{j=1}^{n} x_j^2 \frac{1}{n} = \frac{1}{n} \sum_{j=1}^{n} x_j^2; \quad Var(X) = \frac{\sum_{j=1}^{n} x_j^2}{n} - \left(\frac{\sum_{j=1}^{n} x_j}{n}\right)$$

Example 2.66. X is the $B(p)$ RV, $X \in \{0, 1\}$

$$\mathbb{E}[X^2] = 0 \cdot (1 - p) + p; \quad Var(x) = p - p^2$$

Example 2.67. X is a $B(n, p)$ RV.

$$\mathbb{E}[X^2] = \sum_{k=0}^{n} k^2 \binom{n}{k} p^k (1 - p)^{n-k} = np \sum_{k=1}^{n} k \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} (1 - p)^{n-k}$$

now set $l = k - 1$

$$\mathbb{E}[X^2] = np\sum_{l=0}^{n-1}(l+1)\binom{n-1}{l}p^l(1-p)^{n-1-l}$$

if Y is $B(n-1,p)$ RV then $\mathbb{E}[Y+1] = \sum_{l=0}^{n-1}(l+1)\binom{n-1}{l}p^l(1-p)^{n-1-l} = \mathbb{E}[Y]+1$ so we have

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

and we have

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

Example 2.68. X is $P(\lambda)$ RV.

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

$$= \lambda\sum_{l=0}^{\infty}(l+1)\frac{\lambda^l}{(l)!}e^{-\lambda} = \lambda(\lambda+1)$$

and

$$Var(X) = \lambda(\lambda+1) - \lambda^2$$

EXERCISE: Compute the third, fourth moment for $B(p), B(n,p), P(\lambda)$.

Definition 2.69. Let X be a RV. The Moment Generating Function (MGF) of X is defined as

$$M(s) = \mathbb{E}[e^{sX}]$$

if $\mathbb{E}[e^{sX}]$ exists for s in some neighbourhood near 0, i.e. $\exists s_0 > 0$ such that $\mathbb{E}[e^{sX}] < \infty$ for every $s \in (-s_0, s_0)$.
$M(0) = 1$.

Remark 2.70. Is $X \in \{x_n : n \in \mathbb{N}\}$ then

$$M(s) = \mathbb{E}[e^{sX}] = \sum_n e^{sx_n}\mathbb{P}(X = x_n)$$

Proposition 2.71. *If $M(s)$ exists for $s \in (-s_0, s_0)$ for some $s_0 > 0$, then $\mathbb{E}[X^k]$ exists $\forall k \in \{0,1,2,...\}$ and*

$$M^{(k)}(0) = \frac{d^k}{ds^k}M(s)|_{s=0}$$

exists and

$$\mathbb{E}[X^k] = M^{(k)}(0)$$

Proof. $\forall k \in \{0, 1, 2, ...\}; \forall s \in (-s_0, s_0), s \neq 0$ there exists $C_k > 0$ such that $|y_k| \leq C_k e^{|sy|}$ for all $y \in \mathbb{R}$. (polynomial growth can be controlled by exponential growth).

$$\sum_n |x_n|^k \mathbb{P}(X = x_n) \leq C_k \sum_n e^{|sy|} \mathbb{P}(X = x_n) \leq C_k \sum_n (e^{sx_n} + e^{-sx_n}) \mathbb{P}(X = x_n) = C_k(M(s) + M(-s))$$

Now

$$\frac{d^k}{ds^k} M(s) = \frac{d^k}{ds^k} \left(\sum_n e^{sx_n} \mathbb{P}(X = x_n) \right) |_{s=0}$$

accept the next step

$$= \sum_n \left(\frac{d^k}{ds^k} e^{sx_n} \right) \mathbb{P}(X = x_n)|_{s=0}$$

$$= \sum_n \left(x_n^k e^{sx_n} \right) \mathbb{P}(X = x_n)|_{s=0} = \sum_n x_n^k \mathbb{P}(X = x_n) = \mathbb{E}[X^k]$$

□

Remark 2.72. MGF does not always exist, e.g., $\mathbb{E}[|X|^k] = \infty$ for some k then MGF does not exist.
Even if $\mathbb{E}[X^k]$ exists $\forall k$, MGF does not have to exist.

Remark 2.73. If $\mathbb{E}[X^k]$ exists $\forall k$ and $\exists C > 0$ such that $|\mathbb{E}[X^k]| \leq k! C^k$ for all $k \in \mathbb{N}$ then MGF exists. In this case

$$M(s) = M(0) + M'(0)s + \frac{M''(0)}{2!} s^2 + ...$$

$$M(s) = 1 + \mathbb{E}[X]s + \frac{\mathbb{E}[X^2]}{2!} s^2 + ...$$

Remark 2.74. If $\mathbb{E}[X^2]$ exists then so does $\mathbb{E}[X]$.

Example 2.75. X is a discrete uniform RV, $X \in \{x_1, ..., x_n\}$.

$$M(s) = \mathbb{E}[e^{sX}] = \sum_{i=1}^n e^{sx_i} \mathbb{P}(X = x_i) = \frac{1}{n} \sum_{i=1}^n e^{sx_i}$$

Example 2.76. $X : \Omega \to \{0, 1\}$, a Bernoulli RV, $B(p)$.

$$M(s) = e^0(1 - p) + e^s p = 1 - p + e^s p$$

$$M'(s) = M''(s) = pe^s \quad M'(0) = M''(0) = p = \mathbb{E}[X] = \mathbb{E}[X^2]$$

Example *2.77.* $X : \Omega \to \{0, ..., n\}$ a $B(n, p)$ RV.

$$M(s) = \sum_{k=0}^{n} e^{sk} \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=0}^{n} \binom{n}{k} (pe^s)^k (1-p)^{n-k} = (pe^s + 1 - p)^n$$

$$M'(0) = \mathbb{E}[X] = np \quad M''(0) = \mathbb{E}[X^2] = n(n-1)p^2 + np$$

Example *2.78.* $X : \Omega \to \{0, 1, 2, ...\}$ is a $P(\lambda)$ RV.

$$M(s) = \sum_{k=0}^{\infty} e^{sk} \frac{\lambda^k}{k!} e^{-\lambda} = \sum_{k=0}^{\infty} \frac{(e^s \lambda)^k}{k!} e^{-\lambda} = e^{\lambda(e^s - 1)}$$

$$M'(0) = \lambda = \mathbb{E}[X] \quad M''(0) = \lambda^2 + \lambda = \mathbb{E}[X^2]$$

2.3 Continuous Distributions

Definition *2.79.* Given $(\Omega, \mathcal{S}, \mathbb{P})$ and RV $X : \Omega \to \mathbb{R}$, X is a continuous RV if the distribution of X is continuous, i.e. $\exists f : \mathbb{R} \to [0, \infty)$ such that $\forall x \in \mathbb{R}$

$$\mathbb{P}(X \leq x) = \mathbb{Q}((-\infty, x]) = F(x) = \int_{-\infty}^{x} f(t) dt$$

F is the DF, f is the PDF.

Definition *2.80.* Transformation of a RV:
Consider $g : \mathbb{R} \to \mathbb{R}$, a Borel function. Then $g \circ X$ is a RV.

Example *2.81.* If g is monotonic and g^{-1} exists then if $Y = g \circ X$ then the DF F_Y is given by

$$F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(g \circ X \leq y) = \begin{cases} \mathbb{P}(X \leq g^{-1}(y)) : & \text{if } g \uparrow \\ P(X \geq g^{-1}(y)) : & \text{if } g \downarrow \end{cases} \tag{2.8}$$

Example *2.82.* If $g : \mathbb{R} \to \mathbb{R}$ is monotonic, differentiable, and g^{-1} exists then $g' > 0$ or $g' < 0$. If $g' > 0$ then

$$F_Y(y) = F_X(g^{-1}(y)) = \int_{-\infty}^{g^{-1}(y)} f_X(t) dt = \int_{g(-\infty)}^{y} f_X(g^{-1}(s)) \frac{ds}{g'(g^{-1}(s))}$$

$$F_Y(y) = \int_{g(-\infty)}^{y} \frac{f_X(g^{-1}(s))}{g'(g^{-1}(s))} ds$$

Therefore Y is a continuous RV and the PDF of Y is given by

$$f_Y(y) = \frac{f_X(g^{-1}(y))}{g'(g^{-1}(y))}$$

where $y \in g(\mathbb{R})$.
When $g' < 0$

$$F_Y(y) = \mathbb{P}(X \geq g^{-1}(y)) = \int_{g^{-1}(y)}^{\infty} f_X(t)dt = \int_{g(\infty)}^{y} f_X(g^{-1}(s))\frac{ds}{-g'(g^{-1}(s)}$$

$$F_Y(y) = \int_{g(\infty)}^{y} -\frac{f_X(g^{-1}(s))}{g'(g^{-1}(s))}ds$$

and the PDF of Y is

$$f_Y(s) = -\frac{f_X(g^{-1}(s))}{g'(g^{-1}(s))}$$

Example 2.83. X is a continuous RV with DF F_X and f_X, assume X is a uniform RV on $(a, b), a < b$.

$$F_X(y) = \begin{cases} 0: & \text{if } x \leq a \\ \frac{x-a}{b-a}: & \text{if } a < x < b \\ 1: & \text{if } x \geq b \end{cases} \tag{2.9}$$

For $Y = e^X$ we then have

$$F_Y(y) = \mathbb{P}(e^X \leq y) = \mathbb{P}(X \leq \log y) = \begin{cases} 0: & \text{if } y \leq e^a \\ \frac{\log y - a}{b-a}: & \text{if } e^a < y < e^b \\ 1: & \text{if } y \geq e^b \end{cases} \tag{2.10}$$

$$f_Y(y) = \frac{f_X(g^{-1}(y))}{g'(g^{-1}(y))} = \frac{f_X(\log y)}{y} = \begin{cases} \frac{1}{y(b-a)}: & \text{if } e^a < y < e^b \\ 0: & \text{otherwise} \end{cases} \tag{2.11}$$

Example 2.84. Consider the same X and now $Y = \frac{1}{X-a}$. We have $g(x) = \frac{1}{x-a} \implies g'|_{(a,b)} < 0$.

$$F_Y(y) = \mathbb{P}\left(\frac{1}{X-a} \leq y\right) = \mathbb{P}\left(X \geq a + \frac{1}{y}\right) = \int_{a+\frac{1}{y}}^{\infty} f_X(t)dt$$

$$= \int_0^y \frac{f_X\left(\frac{1}{s} + a\right)}{s^2}ds$$

Then we have

$$f_Y(y) = \begin{cases} \frac{f_X\left(\frac{1}{s}+a\right)}{s^2} = \frac{1}{s^2(b-a)}: & \text{if } \frac{1}{b-a} < y \\ 0: & \text{otherwise} \end{cases} \tag{2.12}$$

Remark 2.85. In general, if g is not monotonic, then you can rely on the definition and treat different g differently, i.e., there is no standard procedure, you have to see what g you have. For example

$$Y = \left(X - \frac{a+b}{2}\right)^2 \implies g = \left(x - \frac{a+b}{2}\right)^2$$

so g is not monotonic.

$$F_Y(y) = \mathbb{P}\left(\left(X - \frac{a+b}{2}\right)^2 \leq y\right) = \mathbb{P}\left(-\sqrt{y} + \frac{a+b}{2} \leq X \leq \sqrt{y} + \frac{a+b}{2}\right)$$

$$F_Y(y) = \int_{-\sqrt{y}+\frac{a+b}{2}}^{\sqrt{y}+\frac{a+b}{2}} f_X(t)dt = \begin{cases} 1 : & \text{if } y > \frac{(b-a)^2}{4} \\ \frac{2\sqrt{y}}{b-a} : & \text{if } 0 < y < \frac{(b-a)^2}{4} \\ 0 : & \text{if } y < 0 \end{cases} \tag{2.13}$$

So Y is a continuous RV with PDF

$$f_Y(y) = \begin{cases} 0 : & \text{if } y > \frac{(b-a)^2}{4} \\ \frac{1}{(b-a)\sqrt{y}} : & \text{if } 0 < y < \frac{(b-a)^2}{4} \\ 0 : & \text{if } y < 0 \end{cases} \tag{2.14}$$

2.3.1 Expectation of a Continuous Random Variable

Definition 2.86. Let X be a continuous RV with PDF f_X. If $\int_{\mathbb{R}} |x| f_X(x)dx < \infty$ exists then $\mathbb{E}[x]$ exists and

$$\mathbb{E}[X] = \int_{\mathbb{R}} x f_X(x)dx$$

Remark 2.87. It is possible that

$$\int_{\mathbb{R}} x f_X(x)dx < \infty$$

but

$$\int_{\mathbb{R}} |x| f_X(x)dx = \infty$$

e.g. if $f_x(x) = \frac{1}{\pi(1+x^2)}$ then

$$\int_{\mathbb{R}} \frac{1}{\pi(1+x^2)}dx = 1; \quad \int_{\mathbb{R}} \frac{x}{\pi(1+x^2)}dx = 0; \quad \int_{\mathbb{R}} \frac{|x|}{\pi(1+x^2)}dx \simeq 2\int_0^\infty \frac{1}{x}dx = \infty$$

Proposition 2.88. *For $Y = g \circ X$, $\mathbb{E}[Y]$ exists if and only if*

$$\int_{\mathbb{R}} |g(x)| f_X(x) dx < \infty$$

In this case $\mathbb{E}[Y] = \mathbb{E}[g \circ X] = \int_{\mathbb{R}} g(x) f_X(x) dx$.

Proof. (Only the case that g is increasing and differentiable with $g' > 0$ on $\mathbb{R}$)
We know that in this case $Y = g \circ X$ is a continuous RV with PDF

$$f_Y(s) = \frac{f_X(g^{-1}(s))}{g'(g^{-1}(s))}$$

Now we want

$$\int_{\mathbb{R}} |y| f_Y(y) dy = \int_{\mathbb{R}} |g(x)| \frac{f_X(g^{-1}(g(x)))}{g'(g^{-1}(g(x)))} g'(x) dx = \int_{\mathbb{R}} |g(x)| f_X(x) dx$$

To compute the $\mathbb{E}[Y]$ we have

$$\int_{\mathbb{R}} y f_Y(y) dy = \int_{\mathbb{R}} g(x) \frac{f_X(g^{-1}(g(x)))}{g'(g^{-1}(g(x)))} g'(x) dx = \int_{\mathbb{R}} g(x) f_X(x) dx$$

□

Definition 2.89. If $X : \Omega \to \mathbb{R}$ is a continuous RV with PDF f_X and if $\mathbb{E}[X^k]$ exists for $k \in \mathbb{N}$ then $\mathbb{E}[X^k]$ is the kth moment. Note: $\mathbb{E}[X^0] = 1$.
$\mathbb{E}[X^k]$ exists if and only if $\int_{\mathbb{R}} |x|^k f_X(x) dx < \infty$ and

$$\mathbb{E}[X^k] = \int_{\mathbb{R}} x^k f_X(x) dx$$

Definition 2.90. If $\mathbb{E}[X^2] < \infty$ then the variance of X is

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

Remark 2.91. $Var(X) \geq 0$ and $Var(X) = 0$ if and only if X is Dirac.

Remark 2.92. Its possible to define $\mathbb{E}[X^\alpha]$ for $\alpha \in \mathbb{R}$.

Definition 2.93. If $\mathbb{E}[e^{sX}] < \infty$ for $s \in (-s_0, s_0)$ for some $s_0 > 0$ then the MGF of X exists and

$$M(s) = \mathbb{E}[e^{sX}]$$

Proposition 2.94. *If MGF $M(s)$ exists in $(-s_0, s_0)$ then $\forall k \in \mathbb{N}$, $\mathbb{E}[X^k]$ exists and $M^{(k)}(0)$ exists and*

$$\mathbb{E}[X^k] = M^{(k)}(0)$$

Proof. $\forall k \in \mathbb{N}, s \in (-s_0, s_0)$ we have $\exists C_k > 0$ such that $|x|^k \leq C_k e^{|sx|}$.

$$\int_{\mathbb{R}} |x|^k f_X(x)dx \leq C_k \int_{\mathbb{R}} e^{|sx|} f_X(x)dx \leq C_k \int_{\mathbb{R}} (e^{sx}+e^{-sx}) f_X(x)dx = C_k(M(s)+M(-s) < \infty$$

So $\mathbb{E}[X^k]$ exists. We also have

$$\frac{d^k}{ds^k} M(s)|_{s=0} = \frac{d^k}{ds^k} \int_{\mathbb{R}} e^{sx} f_X(x)dx|_{s=0} = \int_{\mathbb{R}} \left(\frac{d^k}{ds^k} e^{sx}\right) f_X(x)dx|_{s=0}$$

$$= \int_{\mathbb{R}} e^{sx} x^k f_X(x)dx|_{s=0} = \int_{\mathbb{R}} x^k f_X(x)dx = \mathbb{E}[X^k]$$

□

Example 2.95. X is a uniform RV on (a, b), i.e. the distribution of X is $U((a, b))$ with DF, PDF F_X, f_X.

$$\mathbb{E}[X^k] = \int_{\mathbb{R}} x^k f_X(x)dx = \int_a^b \frac{x^k}{b-a} dx = \frac{b^{k+1} - a^{k+1}}{(b-a)(k+1)}$$

Note that because $\mathbb{P}(a < X < b) = 1$, $\mathbb{E}[X^k], \mathbb{E}[e^{sX}]$ exist.

$$\mathbb{E}[X] = \frac{a+b}{2} \quad \mathbb{E}[X^2] = \frac{a^2+ab+b^2}{3} \quad Var(X) = \frac{(b-a)^2}{12}$$

$$M(s) = \int_a^b \frac{e^{sx}}{b-a} dx = \frac{e^{bs} - e^{as}}{(b-a)s}$$

if $s \neq 0, M(0) = 1$.

Example 2.96. $X : \Omega \to [0, \infty)$ is an exponential RV with parameter $\lambda > 0$ if the DF F_X is given by

$$F_X(x) = \begin{cases} 0 : & \text{if } x \leq 0 \\ 1 - e^{-\lambda x} : & \text{otherwise} \end{cases} \tag{2.15}$$

and PDF

$$f_X(x) = \begin{cases} 0 : & \text{if } x \leq 0 \\ \lambda e^{-\lambda x} : & \text{otherwise} \end{cases} \tag{2.16}$$

We have

$$\mathbb{E}[X] = \int_0^\infty x\lambda e^{-\lambda x} dx = -\int_0^\infty x d(e^{-\lambda x}) = -xe^{-\lambda x}|_0^\infty + \int_0^\infty e^{-\lambda x} dx = \frac{1}{\lambda} e^{-\lambda x}|_\infty^0 = \frac{1}{\lambda}$$

$$\mathbb{E}[X^2] = \int_0^\infty x^2 \lambda e^{-\lambda x} dx = -\int_0^\infty x^2 d(e^{-\lambda x}) = -x^2 e^{-\lambda x}|_0^\infty + \int_0^\infty 2xe^{-\lambda x} dx = \frac{2}{\lambda} \mathbb{E}[X] = \frac{2}{\lambda^2}$$

$$Var(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{1}{\lambda^2}$$
$$M(s) = \int_0^\infty e^{sx}\lambda e^{-\lambda x}dx = \lambda\int_0^\infty e^{-(\lambda-s)x}dx$$
$$M(s) = \begin{cases} \frac{\lambda}{\lambda-s} : & \text{if } s < \lambda \\ \infty : & \text{otherwise} \end{cases} \tag{2.17}$$

So we have $M(s)$ exists for $s \in (-\infty, \lambda)$ with $M(s) = \frac{\lambda}{\lambda-s}$

Remark 2.97. MGF has a 1-1 correspondence to the distribution.

Exercise: Prove that $\mathbb{E}[X^k] = \frac{k!}{\lambda^k}$.

Example 2.98. $X : \Omega \to \mathbb{R}$ is a Gaussian(normal) RV if the PDF is, for $m \in \mathbb{R}, \sigma = \sqrt{\sigma^2}$

$$f_{m,\sigma^2} = \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-m)^2}{2\sigma^2}}$$

i.e. the distribution of X is the Gaussian distribution $N(m, \sigma^2)$.

In particular, if $m = 0, \sigma^2 = 1$ then we say that X has the standard Gaussian distribution, $N(0, 1)$.

If X is $N(0, 1)$ RV then $f_X(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$. Define $Y = \sigma X + m$. Then

$$Y = g \circ X \quad g^{-1}(y) = \frac{y-m}{\sigma} \quad g(x) = \sigma x + m$$

The PDF of Y is

$$f_Y(y) = \frac{f_X\left(\frac{y-m}{\sigma}\right)}{g'\left(\frac{y-m}{\sigma}\right)} = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(y-m)^2}{2\sigma^2}}$$

So Y is the $N(m, \sigma^2)$ RV.

Assume X is $N(0, 1)$ RV, then $\mathbb{E}[X^k]$ exists $\forall k \in \mathbb{N}$ because

$$\int_{\mathbb{R}} |x|^k e^{-\frac{x^2}{2}}dx < \infty$$

Now, since f_X is even, xf_X is odd and we have

$$\mathbb{E}[X] = 0$$
$$Var(X) = \mathbb{E}[X^2] = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} x^2 e^{-\frac{x^2}{2}} = 1$$
$$M(s) = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} e^{sx-\frac{x^2}{2}}dx = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} e^{-\frac{(x-s)^2}{2}+\frac{s^2}{2}}dx = e^{\frac{s^2}{2}}\int_{\mathbb{R}} f_{s,1}(x)dx = e^{\frac{s^2}{2}}$$
$$M''(s) = s^2 e^{\frac{s^2}{2}} + e^{\frac{s^2}{2}}$$
$$Var(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \mathbb{E}[X^2] = M''(0) = 1$$

Exercise: Verify that $\int_{\mathbb{R}} f_{m,\sigma^2}(x)dx = 1$.

Remark 2.99.

$$n!! = \begin{cases} 2 \cdot 4 \cdot 6 \cdot 8 \cdots n : & \text{if } n = 2k \\ 1 \cdot 3 \cdot 5 \cdot 7 \cdots n : & \text{otherwise} \end{cases} \tag{2.18}$$

Exercise: For X a $N(0,1)$ RV, we have

$$\mathbb{E}[X^k] = \begin{cases} 0 : & \text{if } k = 2n+1 \\ (k-1)!! : & \text{otherwise} \end{cases} \tag{2.19}$$

Remark 2.100. If f_X is even, i.e. $f_X(-x) = f_X(x)$, then for k odd we have that if $\mathbb{E}[X^k]$ exists then $\mathbb{E}[X^k] = 0$.

Remark 2.101. If X is $N(0,1)$ RV, $m \in \mathbb{R}, \sigma^2 > 0, \sigma > 0$, then $Y = \sigma X + m$ is $N(m, \sigma^2)$ RV.

$$\mathbb{E}[Y] = \mathbb{E}[\sigma X + m] = \sigma \mathbb{E}[X] + m = m$$

$$\mathbb{E}[Y^2] = \mathbb{E}[(\sigma X + m)^2] + \sigma^2 \mathbb{E}[X^2] + 2\sigma m \mathbb{E}[X] + m^2 = \sigma^2 + m^2$$

$$Var(Y) = \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2 = \sigma^2$$

$$M_Y(s) = \mathbb{E}[e^{sY}] = \mathbb{E}[e^{s\sigma X + sm}] = e^{sm}\mathbb{E}[e^{s\sigma X}] = e^{sm + \frac{s^2\sigma^2}{2}}$$

2.4 Inequalities and Estimates

Remark 2.102. If X is a discrete RV, $X \in \mathbb{N}$ then

$$\mathbb{E}[X] = \sum_{i=1}^{\infty} \mathbb{P}(X \geq i)$$

Note that if the either side diverges, then so does the other.

Proposition 2.103. *If X is a RV and $X \geq 0$ then*

$$\mathbb{E}[X] = \int_0^{\infty} \mathbb{P}(X > t)dt$$

Remark 2.104. If $X \in \mathbb{N}$ then

$$\int_0^{\infty} \mathbb{P}(X > t)dt = \sum_{i=0}^{\infty} \int_i^{i+1} \mathbb{P}(X > t)dt$$

$$= \sum_{i=1}^{\infty} \mathbb{P}(X \geq i)$$

Proof. (Proof only in the case that X is a continuous RV with PDF f_X such that $f_X(x) = 0$ if $x < 0$)

$$\begin{aligned}\int_0^\infty \mathbb{P}(X > t)dt &= \int_0^\infty \int_t^\infty f_X(u)dudt \\ &= \int_0^\infty \int_0^u f_X(u)dtdu \\ &= \int_0^\infty u f_X(u)du \\ &= \mathbb{E}[X]\end{aligned}$$

□

Definition 2.105. For $t > 0, n \in \mathbb{N}, \mathbb{P}(X > T)$ (or $\mathbb{P} \geq n$ or $\mathbb{P}(|X| > t)$) is referred to as the tail probability.
It's important to derive estimates on tail probability.

Corollary 2.106. *For any RV* $X : \Omega \to \mathbb{R}$ *we have*

$$\mathbb{E}[|X|] = \int_0^\infty \mathbb{P}(|X| > t)dt$$

Corollary 2.107. *If* $\exists \delta > 0, C > 0$ *such that* $\mathbb{P}(|X| > t) \leq \frac{C}{t^{1+\delta}}$ *for every* $t \in [0, \infty)$ *then*

$$\mathbb{E}[|X|] = \int_0^\infty \mathbb{P}(|X| > t)dt \leq \int_0^\infty \frac{C}{t^{1+\delta}}dt < \infty$$

Corollary 2.108. *For some* $k \in \mathbb{N}$, *if* $\exists \delta > 0, C > 0$ *such that* $\mathbb{P}(|X| > t) \leq \frac{C}{t^{k+\delta}}$ *for every* $t \in [0, \infty)$ *then*

$$\mathbb{E}[|X|^k] = \int_0^\infty \mathbb{P}(|X|^k > t)dt = \int_0^\infty \mathbb{P}(|X| > t^{\frac{1}{k}})dt \leq \int_0^\infty \frac{C}{t^{1+\frac{\delta}{k}}}dt < \infty$$

Corollary 2.109. *If* $\exists \delta > 0, C > 0$ *such that* $\mathbb{P}(|X| > t) \leq Ce^{-\delta t}$ *for every* $t \in [0, \infty)$ *then* $\forall s \in (-\delta, \delta)$

$$\begin{aligned}\mathbb{E}\left[e^{sX}\right] &= \int_0^\infty \mathbb{P}\left(e^{sX} > t\right)dt \\ &= \int_0^\infty \mathbb{P}(|sX| > \log t)dt \\ &= \int_0^\infty \mathbb{P}\left(|X| > \frac{1}{|s|}\log t\right)dt\end{aligned}$$

$$\leq \int_0^\infty Ce^{-\frac{\delta}{|s|}\log t}dt$$
$$= \int_0^\infty Cte^{-\frac{\delta}{|s|}}dt < \infty$$

Since we have $\frac{\delta}{|s|} > 1$.

Remark 2.110. These are a sufficient conditions for the existence of moments and the generating function, but not a necessary conditions.

Remark 2.111. Now we assume the existence of $\mathbb{E}[X], \mathbb{E}[X^k]$, or the MGF. We derive tail probability estimates.

Proposition 2.112. *Let* $X : \Omega \to \mathbb{R}$ *be a RV,* $h : \mathbb{R} \to \mathbb{R}$ *be Borel with* $h \geq 0$. *Assume that* $\mathbb{E}[h \circ X]$ *exists. Then,* $\forall t > 0$ *we have*

$$\mathbb{P}(h(X) > t) \leq \mathbb{P}(h(X) \geq t) \leq \frac{1}{t}\mathbb{E}[h(X)\chi_{h(X)\geq t}] \leq \frac{\mathbb{E}[h(X)]}{t}$$

Proof. The first inequality is obvious.
The second is true because

$$\mathbb{P}(h(X) \geq t) = \mathbb{E}[\chi_{h(X)\geq t}] \leq \mathbb{E}\left[\frac{h(X)}{t}\chi_{h(X)\geq t}\right]$$

Since $1 \leq \frac{h(X)}{t}$.
The third is true because, since we have $h \geq 0$,

$$h(X)\chi_{h(X)\geq t} \leq h(X)$$

□

Corollary 2.113. *(Markov's Inequality)*
Assume $X : \Omega \to \mathbb{R}$ *is a RV with* $\mathbb{E}[|X|^k] < \infty$ *for some* $k \in \mathbb{N}$. *Then* $\forall t > 0$

$$\mathbb{P}(|X| > t) = \mathbb{P}(|X|^k > t^k) \leq \frac{1}{t^k}\mathbb{E}[|X|^k\chi_{|X|^k>t^k}] \leq \frac{\mathbb{E}[|X|^k]}{t^k}$$

Note we use the above proposition with $h(x) = |x|^k$.

Corollary 2.114. *(Chebyshev's Inequality)*
Assume $X : \Omega \to \mathbb{R}$ *is a RV with* $\mathbb{E}[|X|^2] < \infty$. *Then* $\forall t > 0$

$$\mathbb{P}(|X - \mathbb{E}[X]| > t) = \mathbb{P}(|X - \mathbb{E}[X]|^2 > t^2) \leq \frac{\mathbb{E}[|X - \mathbb{E}[X]|^2]}{t^2} = \frac{1}{t^2}Var(X)$$

Note we use the above proposition with $h(x) = |x - \mathbb{E}[X]|^2$.

Corollary 2.115. *(Chernoff's Bound)*
Assume $X : \Omega \to \mathbb{R}$ is RV with $\mathbb{E}[e^{s|X|}] < \infty$ for some $s > 0$. Then, $\forall t > 0$ we have

$$\mathbb{P}(|X| > t) = \mathbb{P}(e^{s|X|} > e^{st}) \leq \frac{\mathbb{E}[e^{s|X|}]}{e^{st}} = e^{-st}\mathbb{E}[e^{s|X|}]$$

Note we use the above proposition with $h(x) = e^{s|x|}$.

Example 2.116. X is $N(0,1)$ RV. For $c \in (0, \frac{1}{2})$

$$\mathbb{E}[e^{cX^2}] = \sqrt{\frac{1}{1-2c}}$$

$$\mathbb{P}(|X| > t) = \mathbb{P}(e^{cX^2} > e^{ct^2}) \leq \frac{1}{e^{ct^2}}\mathbb{E}[e^{cX^2}] = e^{-ct^2}\sqrt{\frac{1}{1-2c}}$$

This is super exponential decay.

3 Multiple Random Variables: Correlation, Independence, Conditioning

Definition 3.1. Given a probability space $(\Omega, \mathcal{S}, \mathbb{P})$and two functions $X : \Omega \to \mathbb{R}$ and $Y : \Omega \to \mathbb{R}$ then $(X, Y) : \Omega \to \mathbb{R}^2$ i.e. $\forall \omega \in \Omega, (X, Y)(\omega) = (X(\omega), Y(\omega)) \in \mathbb{R}^2$ is a bivariate RV if

$$\forall D \in \mathcal{B}(\mathbb{R}^2) := \sigma(\{\text{all the open sets of } \mathbb{R}^2\}) = \sigma(\{(a,b) \times (c,d) : a < b, c < d \in \mathbb{R}\})$$

we have

$$(X, Y)^{-1}(D) \in \mathcal{S}$$

i.e. $\{\omega \in \Omega : (X(\omega), Y(\omega)) \in D\} \in \mathcal{S}$.
e.g. $D = (a, b) \times (c, d)$

$$(X, Y)^{-1}(D) = \{\omega \in \Omega : X(\omega) \in (a, b), Y(\omega) \in (c, d)\} = \{a < X < b\} \cap \{c < Y < d\} \in \mathcal{S}$$

Proposition 3.2. *$(X, Y) : \Omega \to \mathbb{R}^2$ is a RV if and only if $\forall x, y \in \mathbb{R}, \{X \leq x, Y \leq y\} = \{X \leq x\} \cap \{Y \leq y\} \in \mathcal{S}$.*

Proof. ($\Longrightarrow$) Because if $D = (-\infty, x] \times (-\infty, y] \in \mathcal{B}(\mathbb{R}^2)$ then $\{X \leq x, Y \leq y\} = (X, Y)^{-1}(D) \in \mathcal{S}$.
($\Longleftarrow$) Use the fact that $\mathcal{B}(\mathbb{R}^2) = \sigma(\{(-\infty, x] \times (-\infty, y] : x, y \in \mathbb{R}\})$. □

Remark 3.3. $\mathcal{B}(\mathbb{R}^2) = \mathcal{B}(\mathbb{R}) \otimes \mathcal{B}(\mathbb{R}) \neq \mathcal{B}(\mathbb{R}) \times \mathcal{B}(\mathbb{R})$.
We have the open unit disc is in $\mathcal{B}(\mathbb{R}^2)$ but not $\mathcal{B}(\mathbb{R}) \times \mathcal{B}(\mathbb{R})$.

Proposition 3.4. *$(X,Y):\Omega\to\mathbb{R}^2$ is a RV if and only if $X:\Omega\to\mathbb{R}$ is a RV, $Y:\Omega\to\mathbb{R}$ is a RV.*

Proof. ($\Longleftarrow$) $\forall x,y\in\mathbb{R};\{X\le x\}\in\mathcal{S},\{Y\le y\}\in\mathcal{S}$. By the previous proposition, (X,Y) is a RV.
($\Longrightarrow$) $\forall B\in\mathcal{B}(\mathbb{R})$

$$X^{-1}(B)=\{\omega\in\Omega:X(\omega)\in B,Y(\omega)\in\mathbb{R}\}=(X,Y)^{-1}(B\times\mathbb{R})\in\mathcal{S}$$

So X is a RV and similarly, so is Y. □

Definition 3.5. Given a probability space $(\Omega,\mathcal{S},\mathbb{P})$ and n functions $X_i:\Omega\to\mathbb{R}$ then $(X_1,...,X_n):\Omega\to\mathbb{R}^n$ i.e. $\forall\omega\in\Omega,(X_1,...,X_n)(\omega)=(X_1(\omega),...,X_n(\omega))\in\mathbb{R}^2$ is a multivariate RV if

$$\forall D\in\mathcal{B}(\mathbb{R}^n),(X_1,...,X_n)^{-1}(D)\in\mathcal{S}$$

$$\iff\forall x_i\in\mathbb{R},\{X_1\le x_1,...,X_n\le x_n\}\in\mathcal{S}$$

$$\iff X_i:\Omega\to\mathbb{R}\text{ is a RV }\forall i=1,...,n$$

Remark 3.6. Everything we do with bivariate RVs we can extend to multivariate RVs.

3.1 The Distribution of (X,Y)

Definition 3.7. Given the RV (X,Y), the joint distribution function of (X,Y) is defined by $F_{(X,Y)}:\mathbb{R}^2\to[0,1];(x,y)\mapsto\mathbb{P}(X\le x,Y\le y)$.

3.1.1 Facts about the joint DF

Remark 3.8. $0\le F_{(X,Y)}\le 1$. For any fixed $x\in\mathbb{R}$ we have

$$\{X\le x,Y\le y\}\uparrow_{y\to\infty}\{X\le x,Y<\infty\}=\{X\le x\}$$

$$F_{(X,Y)}=\mathbb{P}(X\le x,Y\le y)\uparrow_{y\to\infty}\mathbb{P}(X\le x)=F_X(x)$$

So we have the marginal DF of X

$$F_X(x)=\lim_{y\to\infty}F_{(X,Y)}$$

Similarly for and fixed $y\in\mathbb{R}$, we have the marginal DF of Y

$$F_Y(y)=\lim_{x\to\infty}F_{(X,Y)}$$

and also

$$\lim_{x,y\to\infty}F_{(X,Y)}=1$$

Remark 3.9. For each fixed $x \in \mathbb{R}$, $\lim_{y\to-\infty} F(x,y) = \mathbb{P}(X \le x, Y < -\infty) = 0$
For each fixed $y \in \mathbb{R}$, $\lim_{x\to-\infty} F(x,y) = \mathbb{P}(Y \le y, X < -\infty) = 0$

Remark 3.10. For each fixed $x \in \mathbb{R}$, $F(x,y)$ is nondecreasing in y and right continuous in y because $(-\infty, x] \times (-\infty, y+\epsilon] \downarrow (-\infty, x] \times (-\infty, y]$ as $\epsilon \downarrow 0$.

$$\implies \lim_{\epsilon \downarrow 0^+} F(x, y+\epsilon) = F(x,y)$$

Similarly for fixed $y \in \mathbb{R}$, $F(x,y)$ is nondecreasing and right continuous in x.

Remark 3.11. If $(X,Y) : \Omega \to \mathbb{R}^2$ is a discrete RV, i.e. it takes countable values in $\mathbb{R}^2$, then X, Y are a discrete RVs, say

$$X \in \{x_i : i \in \mathbb{N}\}, Y \in \{y_j : j \in \mathbb{N}\} \implies (X,Y) \in \{(x_i, y_j) : i, j \in \mathbb{N}\}$$

The function $f_{(X,Y)}(x_i, y_j) = \mathbb{P}(X = x_i, Y = y_j)$ is the joint PMF of (X,Y).
Fix x_i. Then the marginal PMF of X is

$$\sum_{j=1}^{\infty} f_{(X,Y)}(x_i, y_j) = \sum_{j=1}^{\infty} \mathbb{P}(X = x_i, Y = y_j) = \mathbb{P}(X = x_i) = f_X(x_i)$$

This is similar for each y_j.

Remark 3.12. $(X,Y) : \Omega \to \mathbb{R}^2$ is a continuous RV if there exists a nonnegative function $f_{(X,Y)} : \mathbb{R}^2 \to [0, \infty)$ such that

$$\forall (x,y) \in \mathbb{R}^2 \quad F_{(X,Y)}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{(X,Y)}(u,v)\,dv\,du$$

$f_{(X,Y)}$ is the joint PDF of (X,Y). In this case the distribution of (X,Y) is a continuous distribution on $\mathbb{R}^2$.

For fixed y

$$F_Y(y) = \lim_{x\to\infty} F_{(X,Y)}(x,y) = \lim_{x\to\infty} \int_{-\infty}^{x} \int_{-\infty}^{y} F_{(X,Y)}(u,v)\,dv\,du = \int_{-\infty}^{y} \left(\int_{\mathbb{R}} f_{(X,Y)}(u,v)\,du \right) dv$$

We denote $\int_{\mathbb{R}} f_{(X,Y)}(u,v)\,du = f_Y(y)$ now

$$F_Y(y) = \int_{-\infty}^{y} f_Y(v)\,dv$$

so Y is a continuous RV with marginal PDF $f_Y(y) = \int_{\mathbb{R}} f_{(X,Y)}(u,v)\,du$.

Similarly for fixed x,

$$f_X(x) = \int_{\mathbb{R}} f_{(X,Y)}(u,v)\,dv; \quad F_X(x) = \int_{\mathbb{R}} f_X(u)\,du$$

Example 3.13. $(X, Y) : \Omega \to \mathbb{R}^2$ has the uniform distribution on $[0,1]^2$. (X, Y) is a continuous RV with joint PDF

$$f_{(X,Y)}(x, y) = \begin{cases} 1: & \text{if } (x, y) \in [0,1]^2 \\ 0: & \text{otherwise} \end{cases} \tag{3.1}$$

$$F_{(X,Y)}(x, y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f_{(X,Y)}(u, v) du dv = \begin{cases} 0: & \text{if } (x, y) \in (\infty, 0]^2 \\ xy: & \text{if } (x, y) \in [0,1]^2 \\ x: & \text{if } (x, y) \in [0,1] \times [1, \infty) = F_X(x)F_Y(y) \\ y: & \text{if } (x, y) \in [1, \infty) \times [0,1] \\ 1: & \text{if } (x, y) \in [1, \infty)^2 \end{cases} \tag{3.2}$$

$$f_Y(y) = \begin{cases} 1: & \text{if } y \in [0,1] \\ 0: & \text{otherwise} \end{cases} \tag{3.3}$$

$$f_X(x) = \begin{cases} 1: & \text{if } x \in [0,1] \\ 0: & \text{otherwise} \end{cases} \tag{3.4}$$

Both f_X and f_Y are the PDF of $U([0,1])$.

Remark 3.14. $f_{(X,Y)}(x, y) = f_X(x) \cdot f_Y(y)$ then X, Y are independent.

Remark 3.15. PICTURE

3.1.2 Transformation of (X, Y)

Proposition 3.16. *Let $(X, Y) : \Omega \to \mathbb{R}^2$ be a RV and $g : \mathbb{R}^2 \to \mathbb{R}$ be Borel, i.e. $\forall B \in \mathcal{B}(\mathbb{R}), g^{-1}(B) \in \mathcal{B}(\mathbb{R}^2)$. Then $g \circ (X, Y) = g(X, Y) : \Omega \to \mathbb{R}$ is a RV.*

Definition 3.17. Discrete Distribution of $g(X, Y)$
$(X, Y) \in \{(x_i, y_j) : i, j \in \mathbb{N}\}; g(X, Y)$ a discrete RV with $g(X, Y) \in \{z_l : l \in \mathbb{N}\}$. We have PMF

$$\mathbb{P}(g(X, Y) = z_l) = \sum_{(i,j):g(x_i,y_j)=z_l} \mathbb{P}(X = x_i, Y = y_j)$$

Proposition 3.18.

$$\mathbb{E}[g(X, Y)] \text{ exists} \iff \sum_{i,j=1}^{\infty} |g(x_i, y_j)| \mathbb{P}(X = x_i, Y = y_j) < \infty$$

In this case

$$\mathbb{E}[g(X, Y)] = \sum_{i,j=1}^{\infty} g(x_i, y_j) \mathbb{P}(X = x_i, Y = y_j) < \infty$$

Corollary 3.19.

$$a, b \in \mathbb{R}, g : \mathbb{R}^2 \to \mathbb{R}, g(x, y) = ax + by \implies \mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$$

Definition 3.20. Continuous Distribution of $g(X, Y)$
$(X, Y) : \Omega \to \mathbb{R}^2$ is continuous with joint PDF $f_{(X,Y)}(x, y)$, $g : \mathbb{R}^2 \to \mathbb{R}$ Borel. We have

$$\mathbb{P}(g(X, Y) \leq z) = \int\int_{(x,y):g(x,y)\leq z} f_{(X,Y)}(x, y)dxdy$$

Proposition 3.21.

$$\mathbb{E}[g(X, Y)] \textit{ exists} \iff \int_{\mathbb{R}}\int_{\mathbb{R}} |g(x, y)| f_{(X,Y)}(x, y)dxdy < \infty$$

In this case

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

Example 3.22. (X, Y) is $U([0, 1]^2)$, $g : \mathbb{R}^2 \to \mathbb{R}; (x, y) \mapsto x + y$.

$$\mathbb{P}(X + Y \leq z) = \int\int_{(x,y):x+y\leq z} \chi_{[0,1]^2}(x, y)dxdy$$

The domain of integration is then the intersection of the square $[0, 1]^2$ with the area below the line $y = z - x$. This is

$$F_Z(z) = \begin{cases} 0 : & \text{if } z \leq 0 \\ \frac{1}{2}z^2 : & \text{if } 0 \leq z \leq 1 \\ 1 - \frac{(2-z)^2}{2} : & \text{if } 1 \leq z \leq 2 \\ 1 : & \text{if } 2 < z \end{cases} \tag{3.5}$$

Also we have

$$f_Z(z) = \begin{cases} 0 : & \text{if } z \leq 0, z \geq 2 \\ z : & \text{if } 0 \leq z \leq 1 \\ 2 - z : & \text{if } 1 \leq z \leq 2 \end{cases} \tag{3.6}$$

$$\mathbb{E}[Z^2] = \int_{\mathbb{R}} z^2 f_Z(z)dz = \int_0^1 z^3 dz + \int_1^2 z^2(2 - z)dz = \frac{7}{6}$$

What if I want to study $U = X + Y, V = X - Y$ simultaneously?

Theorem 3.23. *$(X,Y):\Omega\to\mathbb{R}^2$ is a continuous RV with PDF $f_{(X,Y)}(x,y)$, $f_1,f_2:\mathbb{R}^2\to\mathbb{R}$ Borel functions with*

$$\begin{cases} U = f_1(x,y) \\ V = f_2(x,y) \end{cases} \tag{3.7}$$

If this can be solved uniquely for

$$\begin{cases} x = g_1(u,v) \\ y = g_2(u,v) \end{cases} \tag{3.8}$$

and g_1, g_2 have continuous partial derivatives, with the Jacobian matrix being invertible, i.e.

$$\det \frac{\partial(g_1,g_2)}{\partial(u,v)} = \det \begin{pmatrix} \frac{\partial g_1}{\partial u} & \frac{\partial g_1}{\partial v} \\ \frac{\partial g_2}{\partial u} & \frac{\partial g_2}{\partial v} \end{pmatrix} \neq 0$$

then (U,V) is a continuous RV and the PDF is

$$f_{(U,V)}(u,v) = f_{(X,Y)}(g_1(u,v), g_2(u,v)) \left| \det \frac{\partial(g_1,g_2)}{\partial(u,v)} \right|$$

Example 3.24. We have $f_1 = x+y, f_2 = x-y$

$$\begin{cases} u = x+y \\ v = x-y \end{cases} \implies \begin{cases} x = \frac{u+v}{2} \\ y = \frac{u-v}{2} \end{cases} \tag{3.9}$$

$$\frac{\partial(g_1,g_2)}{\partial(u,v)} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

$$f_{(U,V)}(u,v) = f_{(X,Y)}\left(\frac{u+v}{2}, \frac{u-v}{2}\right) \cdot \frac{1}{2}$$

Example 3.25. If (X,Y) is the uniform RV on $[0,1]^2$, and we have $u = x+y, v = x-y$ then

$$f_{(U,V)}(u,v) = \begin{cases} \frac{1}{2}: & \text{if } 0 \le \frac{u+v}{2}, \frac{u-v}{2} \le 1 \\ 0: & \text{otherwise} \end{cases} \tag{3.10}$$

Now we recover the marginal distribution of U.

$$f_U(u) = \begin{cases} 0: & \text{if } 0 > u; 2 < u \\ u: & \text{if } 0 \le u \le 1 \\ 2-u: & \text{if } 1 \le u \le 2 \end{cases} \tag{3.11}$$

Example 3.26.

$$\begin{cases} U = \sqrt{X^2+Y^2} \implies u = f_1 = \sqrt{x^2+y^2} \\ V = \arctan \frac{Y}{X} \implies v = f_2 = \arctan \frac{y}{x} \end{cases} \tag{3.12}$$

Note: we have two problems. First, f_2 is not defined if $x = 0$. Simply view (X, Y) as a RV defined on $\mathbb{R}^2 - \{0\}$. Second, f_1, f_2 does not have a unique solution, as $(-x, -y), (x, y)$ map to the same values of f_1, f_2.

To deal with this, first we will restrict X, Y to $R := \{(x, y) : x > 0\} = \{(u, v) : v \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)\}$. Now

$$\begin{cases} u = \sqrt{x^2+y^2} \\ v = \arctan \frac{y}{x} \end{cases} \implies \begin{cases} x = u\cos v = g_1 \\ y = u \sin v = g_2 \end{cases} \implies \frac{\partial(g_1, g_2)}{\partial(u, v)} = \begin{pmatrix} \cos v & -u\sin v \\ \sin v & u\cos v \end{pmatrix} \tag{3.13}$$

So on R we have $\det \frac{\partial(g_1,g_2)}{\partial(u,v)} = u > 0$ so

$$f_{(U,V)}(u, v) = f_{(X,Y)}(u\cos v, u\sin v) \cdot u$$

Second we will restrict X, Y to $L := \{(x, y) : x < 0\} = \{(u, v) : v \in \left(-\frac{\pi}{2}, \frac{3\pi}{2}\right)\}$. Now

$$\begin{cases} u = \sqrt{x^2+y^2} \\ v = \arctan \frac{y}{x} \end{cases} \implies \begin{cases} x = -u\cos v = g_1 \\ y = -u \sin v = g_2 \end{cases} \implies \begin{vmatrix} -\cos v & u\sin v \\ -\sin v & -u\cos v \end{vmatrix} = u > 0 \tag{3.14}$$

$$f_{(U,V)}(u, v) = f_{(X,Y)}(-u\cos v, -u\sin v) \cdot u$$

$$\mathbb{P}((U, V) \in D) = \mathbb{P}\left(\left(\sqrt{x^2+y^2}, \arctan\frac{y}{x}\right) \in D, x > 0\right) + \mathbb{P}\left(\left(\sqrt{x^2+y^2}, \arctan\frac{y}{x}\right) \in D, x < 0\right)$$

$$= \int\int_{\left(\sqrt{x^2+y^2}, \arctan\frac{y}{x}\right)\in D, x>0} f_{(X,Y)} dxdy + \int\int_{\left(\sqrt{x^2+y^2}, \arctan\frac{y}{x}\right)\in D, x<0} f_{(X,Y)} dxdy$$

$$= \int\int_{(u,v)\in D, v\in\left(-\frac{\pi}{2},\frac{\pi}{2}\right)} u f_{(U,V)}(u\cos v, u\sin v) dudv + \int\int_{(u,v)\in D, v\in\left(-\frac{\pi}{2},\frac{\pi}{2}\right)} u f_{(U,V)}(-u\cos v, -u\sin v) dudv$$

$$= \int\int_{(u,v)\in D, v\in\left(-\frac{\pi}{2},\frac{\pi}{2}\right)} u\left(f_{(U,V)}(u\cos v, u\sin v) dudv + f_{(U,V)}(-u\cos v, -u\sin v)\right) dudv$$

Exercise: Determine the PDF of $U = \sqrt{X^2+Y^2}, V = \arctan\frac{Y}{X}$ for (X, Y) being the uniform RV on $(0, 1)^2$ as well as the marginal PDF of $f_U(u)$.

Remark 3.27. The joint DF/PMF/PDF give rise to marginal DF/PMF/PDF.
It is possible to have two different joint distributions that produce the same set of marginal distributions.
Thus, given the marginal distribution alone, in general one can't determine the joint distributions.
However, this can be done when RVs are independent.

3.1.3 Independence of Random Variables

Definition 3.28. Assume $X_1, ..., X_n$ are RVs on Ω. $X_1, ..., X_n$ are independent if and only if

$$\forall x_1, ..., x_n \in \mathbb{R}, F_{(X_1,...,X_n)}(x_1, ..., x_n) = \prod_{i=1}^{n} F_{X_i}(x_i)$$

if and only if

$$\forall x_1, ..., x_n \in \mathbb{R}, \mathbb{P}\left(\bigcap_{i=1}^{n}\{X_i \leq x_i\}\right) = \prod_{i=1}^{n} \mathbb{P}(X_i \leq x_i)$$

if and only if

$$\forall B_1, ..., B_n \in \mathcal{B}(\mathbb{R}), \mathbb{P}\left(\bigcap_{i=1}^{n}\{X_i \in B_i\}\right) = \prod_{i=1}^{n} \mathbb{P}(X_i \in B_i) = \mathbb{P}\left((X_1, ..., X_n)^{-1}(B_1 \times \cdots \times B_n)\right)$$

Definition 3.29. Let $\{X_i : i \in \mathbb{N}\}$ be a sequence of RVs on Ω, then $\{X_i : i \in \mathbb{N}\}$ is an independent family of RVs (or X_i's are independent) if any finite sub collection is independent.

Proposition 3.30. *Let X, Y be independent RVs on Ω; $f, g : \mathbb{R} \to \mathbb{R}$ be Borel. Then $f(X), g(X)$ are also independent.*

Proof. $\forall A, B \in \mathcal{B}(\mathbb{R})$ we have

$$\mathbb{P}(f(X) \in A, g(Y) \in B) = \mathbb{P}(X \in f^{-1}(A), Y \in g^{-1}(B))$$

$$= \mathbb{P}(X \in f^{-1}(A)) \cdot \mathbb{P}(Y \in g^{-1}(B)) = \mathbb{P}(f(X) \in A) \cdot \mathbb{P}(g(Y) \in B)$$

□

Proposition 3.31. *Let X, Y be independent; f, g Borel. If $\mathbb{E}[X], \mathbb{E}[Y]$ exist then $\mathbb{E}[f(X) \cdot g(Y)]$ exists and*

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

e.g.

$$(1) \quad \mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$$

$$(2) \quad \forall A \in \mathcal{B}(\mathbb{R}), \mathbb{E}\left[X\chi_{Y \in A}\right] = \mathbb{E}[X]\mathbb{P}(Y \in A) \implies \chi_{Y \in A} = \chi_A(Y)$$

$$(3) \quad s, t \in \mathbb{R}, M_{(X,Y)}(s, t) = \mathbb{E}\left[e^{sX+tY}\right] = \mathbb{E}\left[e^{sX}\right]\mathbb{E}\left[e^{tY}\right] = M_X(s)M_Y(t)$$

Theorem 3.32. *We have X, Y are RVs on Ω.*

- *if X, Y are discrete, then X, Y are independent $\iff$ the PMF*

$$\mathbb{P}(X = x_i, Y = y_j) = \mathbb{P}(X = x_i)\mathbb{P}(Y = y_j)$$

The forward direction is obvious, for the reverse

$$\forall x, y \in \mathbb{R}, \mathbb{P}(X \leq x, Y \leq y) = \sum_{i \in \mathbb{N}} \sum_{j \in \mathbb{N}} \mathbb{P}(X = x_i)\mathbb{P}(Y \leq y_j) = \mathbb{P}(X = x_i)\mathbb{P}(Y = y_j)$$

- *if X, Y are continuous RVs then X, Y are independent $\iff$ the joint PDF $f_{(X,Y)}(x, y) = f_X(x)f_Y(y)$. For the reverse direction, $\forall x, y \in \mathbb{R}$*

$$F_{(X,Y)}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_X(u)f_Y(v)dvdu = \int_{-\infty}^{x} f_X(u)du \int_{-\infty}^{y} f_Y(v)dv = F_X(x)F_Y(y)$$

For the forward direction,

$$F_X(x)F_Y(y) = F_{(X,Y)}(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{(X,Y)}(u, v)dvdu$$

$$\implies f_X(x)f_Y(y) = \frac{\partial^2}{\partial x \partial y} F_X(x)F_Y(y) = \frac{\partial^2}{\partial x \partial y} F_{(X,Y)}(x, y) = f_{(X,Y)}(x, y)$$

Example 3.33. $X : \Omega \to (0, \infty)$, an exponential RV with parameter 1; $Y : \Omega \to (0, \infty)$, an exponential RV with parameter 2; X, Y independent. Determine the DF of $Z = \frac{Y}{X}$.

$$f_{(X,Y)}(x, y) = f_X(x)f_Y(y) = \begin{cases} 2e^{-x-2y} : & \text{if } x > 0, y > 0 \\ 0 : & \text{otherwise} \end{cases} \tag{3.15}$$

Now $\forall z \in \mathbb{R}$

$$\mathbb{P}(Z \leq z) = \mathbb{P}\left(\frac{Y}{X} \leq z\right) = \int\int_{\{\frac{y}{x} \leq z\}} f_{(X,Y)}(x, y)dxdy = \int\int_{\{x>0, y>0, y \leq zx\}} 2e^{-x-2y}dxdy$$

$$= \int_0^{\infty} \int_0^{xz} 2e^{-x-2y}dydx = \begin{cases} \frac{2z}{2z+1} : & \text{if } z > 0 \\ 0 : & \text{otherwise} \end{cases} \tag{3.16}$$

This then gives,

$$F_Z(z) = \begin{cases} \frac{2z}{2z+1} = 1 - \frac{1}{2z+1} : & \text{if } z > 0 \\ 0 : & \text{otherwise} \end{cases}; \quad f_Z(z) = \begin{cases} \frac{2}{(2z+1)^2} : & \text{if } z > 0 \\ 0 : & \text{otherwise} \end{cases} \tag{3.17}$$

Definition 3.34. If $\{X_i : i \in \mathbb{N}\}$ is an independent sequence of RVs on Ω, we say that $\{X_i : i \in \mathbb{N}\}$ is IDD (independent and identically distributed) if $\{X_i : i \in \mathbb{N}\}$ is independent and

$$F_{X_1} = F_{X_2} = \cdots = F_{X_i} = \cdots = F$$

3.1.4 Facts about the Sum of Independent Variable

Remark 3.35. If $\{X_i : i \in \mathbb{N}\}$ are IDD Bernoulli RVs then for $S_n = \sum_{i=1}^n X_i$ we have

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

Which is the PMF of $B(n, p)$.

Remark 3.36. For X, Y independent – X a $B(n, p)$ RV on Ω, Y a $B(m, p)$ RV on Ω – we have $X + Y \in \{0, 1, ..., n + m\}$ and

$$\mathbb{P}(X+Y=k) = \sum_{q=0}^{k} \mathbb{P}(X=q, Y=k-q) = \sum_{q=0}^{k} \mathbb{P}(X=q)\mathbb{P}(Y=k-q)$$

$$= \sum_{q=0}^{k} \binom{n}{q} p^q (1-p)^{n-q} \binom{m}{k-q} p^{k-q} (1-p)^{m-k+q}$$

$$= p^k (1-p)^{n+m-k} \sum_{q=0}^{k} \binom{n}{q}\binom{m}{k-q} = p^k (1-p)^{m+n-k} \binom{n+m}{k}$$

This is the PMF of $B(n + m, p)$. Thus, by induction, we can show that for $X_i, i = 1, ..., m$ each $B(n_i, p)$ RVs, we have that $\sum_{i=0}^m X_i$ is a $B\left(\sum_{i=0}^m n_i, p\right)$ RV.

Remark 3.37. For X, Y independent RVs on Ω – where X is $P(\lambda_1)$, Y is $P(\lambda_2)$ – we have

$$\mathbb{P}(X+Y=k) = \sum_{q=0}^{k} \mathbb{P}(X=q)\mathbb{P}(Y=k-q) = \sum_{q=0}^{k} e^{-\lambda_1-\lambda_2} \frac{\lambda_1^q}{q!} \frac{\lambda_2^{k-q}}{(k-q)!}$$

$$= \frac{e^{-\lambda_1-\lambda_2}}{k!} \sum_{q=0}^{k} \frac{\lambda_1^q \lambda_2^{k-q} k!}{q!(k-q)!} = \frac{e^{-\lambda_1-\lambda_2}}{k!} \sum_{q=0}^{k} \binom{k}{q} \lambda_1^q \lambda_2^{k-q} = e^{-\lambda_1-\lambda_2} \frac{(\lambda_1+\lambda_2)^k}{k!}$$

This is the PMF of $P(\lambda_1 + \lambda_2)$. In general, for $X_1, ..., X_m$ – where X_i is a $P(\lambda_i)$ RV – we will have that $\sum_{i=0}^m X_i$ is a $P\left(\sum_{i=0}^m \lambda_i\right)$ RV.

Remark 3.38. The Sum of IDD $U([0, 1])$ RVs is HW.

Remark 3.39. For X, Y independent exponential RVs with parameter $\lambda > 0$, we define $D = \{(x, y) : x + y \leq z; x, y > 0\}$ – i.e. the triangle with vertices $(0, 0), (0, z), (z, 0)$ – and we have

$$F_{X+Y}(z) = \mathbb{P}(X+Y \leq z) = \int\int_D \lambda^2 e^{-\lambda x} e^{-\lambda y} dx dy = \int_0^z \int_0^{z-x} \lambda^2 e^{-\lambda x} e^{-\lambda y} dy dx$$

$$= \int_0^z \lambda e^{-\lambda x} (1 - e^{-\lambda(x-z)}) dx = 1 - e^{-\lambda z} - e^{-\lambda z} \lambda z$$

So we have

$$F_{X+Y}(z) = \begin{cases} 1 - e^{-\lambda z} - e^{-\lambda z}\lambda z : & \text{if } z > 0 \\ 0 : & \text{otherwise} \end{cases}; \quad f_{X+Y}(z) = \begin{cases} \lambda^2 e^{-\lambda z} z : & \text{if } z > 0 \\ 0 : & \text{otherwise} \end{cases} \tag{3.18}$$

Recall from HW3 that the PDF of the $\Gamma(\alpha, \lambda)$ distribution is

$$f_{X+Y}(z) = \begin{cases} \lambda^\alpha e^{-\lambda z} z^{\alpha-1} \cdot \frac{1}{\Gamma(\alpha)} : & \text{if } z > 0 \\ 0 : & \text{otherwise} \end{cases} \tag{3.19}$$

So by induction we will have that the sum of n copies of an IID exponential RV with parameter $\lambda > 0$ is $\Gamma(n, \lambda)$.

Alternatively we could find this using the MGFs for these RVs

$$M_{X_1+\cdots+X_n}(s) = \prod_{i=1}^n M_{X_i}(s) = \prod_{i=1}^n \left(\frac{\lambda}{\lambda - s}\right) = \left(\frac{\lambda}{\lambda - s}\right)^n$$

This is the MGF of $\Gamma(n, \lambda)$.

Remark 3.40. For $\{X_i : i \in \mathbb{N}\}$ IID exponential RVs on $(\Omega, \mathcal{S}, \mathbb{P})$ with parameter $\lambda > 0$ we define

$$S_n = \sum_{i=1}^n X_i; \quad S_0 = 0$$

We also define $\forall t > 0$, $N_t : \Omega \to \mathbb{N}$ by

$$N_t(\omega) = \max\{n \in \mathbb{N} : S_n(\omega) \leq t\}$$

N_t is the number of occurrences by time t. N_t is a RV, $\{N_t = k\} = \{S_k \leq t, S_{k+1} \geq t\} \in \mathcal{S}$. We have

$$f_{(S_k, X_{k+1})}(s, k) = f_{S_k}(s) f_{X_{k+1}}(x) = \lambda^k e^{-\lambda s} s^{k-1} \cdot \frac{1}{\Gamma(k)} \lambda e^{-\lambda x}$$

$$\begin{aligned} \mathbb{P}(N_t = k) &= \mathbb{P}(S_k \leq t, S_{k+1} > t) = \mathbb{P}(S_k \leq t, S_k + X_{k+1} > t) \\ &= \int\int_{\{(s,x):0<s\leq t, s+x>t, x>0\}} \lambda^k e^{-\lambda s} s^{k-1} \cdot \frac{1}{\Gamma(k)} \lambda e^{-\lambda x} ds dx \\ &= \int_0^t \left(\int_{t-s}^\infty \lambda e^{-\lambda x} dx\right) \lambda^k e^{-\lambda s} s^{k-1} \cdot \frac{1}{\Gamma(k)} ds = \frac{(\lambda t)^k e^{-\lambda t}}{k!} \end{aligned}$$

This is the PMF of $P(\lambda t)$.

Remark 3.41. For $X_1, ..., X_n$ independent Gaussian RVs on Ω with distributions $N(m_i, \sigma_i^2), i = 1, ..., n$ we define $S_n = \sum_{i=1}^n X_i$. This has the DF

$$\mathbb{P}(S_n \leq z) = \int \cdots \int \prod_{i=1}^n \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{(x_i - m_i)^2}{2\sigma_i^2}} dx_1 \cdots dx_n$$

This is too complicated, so we look at the MGF

$$M_{X_1 + \cdots + X_n}(s) = \prod_{i=1}^n M_{X_i}(s) = e^{s(m_1 + \cdots + m_n) + \frac{s^2}{2}(\sigma_1^2 + \cdots + \sigma_n^2)}$$

This is the MGF of $N(m_1 + \cdots + m_n, \sigma_1^2 + \cdots + \sigma_n^2)$ so the sum of independent RVs is a Gaussian RV!

Remark 3.42. For $X_1, ..., X_n$ IDD $N(m, \sigma^2)$ RVs, we have $S_n = \sum_{i=1}^n X_i$ is a $N(nm, n\sigma^2)$ RV. Then we have that

$$\frac{S_n - nm}{\sqrt{n}\sigma}$$

is a $N(0,1)$ RV. (Recall that for X a $N(0,1)$ RV, we have $Y = \sigma X + m$ is a $N(m, \sigma^2)$ RV)

3.2 Covariance

Definition 3.43. Let X and Y be two RVs on $(\Omega, \mathcal{S}, \mathbb{P})$ and $\mathbb{E}[X^2], \mathbb{E}[Y^2] < \infty$. Then the covariance between X and Y is

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

Remark 3.44. $Cov(X, X) = Var(X)$

Proposition 3.45. $|Cov(X, Y)| \leq \sqrt{Var(X)Var(Y)}$

Proof. This is a consequence of the Caucy-Schwarz inequality, i.e. $|\mathbb{E}[XY]| \leq \sqrt{\mathbb{E}[X^2]\mathbb{E}[Y^2]}$.

$$|Cov(X, Y)| = |\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]| \leq \sqrt{(\mathbb{E}[X - \mathbb{E}[X]])^2(\mathbb{E}[Y - \mathbb{E}[Y]])^2} = \sqrt{Var(X)Var(Y)}$$

□

Definition 3.46. Let X and Y be two RVs on $(\Omega, \mathcal{S}, \mathbb{P})$ and $\mathbb{E}[X^2], \mathbb{E}[Y^2] < \infty$. Then the correlation between X and Y is

$$Corr(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}}$$

Remark 3.47. It follows from the previous proposition that $|Corr(X, Y)| \leq 1$.

Remark 3.48. $Corr(X,Y)$ resembles how much the distribution of (X,Y) resembles a line.

Remark 3.49. We have the following facts:

- $|Corr(X,Y)| = 1 \iff \exists c > 0$ such that $Y = cX$
- $Corr(X,Y) \in [-1,1]$
- $Corr(X,Y)$ is only defined if $Var(X), Var(Y) > 0$

Remark 3.50. $Cov(X,Y)$ does not determine the distribution of (X,Y). However, in some special cases – e.g., if (X,Y) is a jointly Gaussian RV – then $\mathbb{E}[X], \mathbb{E}[Y], Var(X), Var(Y), Cov(X,Y)$ completely determine the joint PDF.

Proposition 3.51. *If X,Y are independent, then $Cov(X,Y) = 0$.*

Proof.

$$Cov(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 0$$

□

Corollary 3.52. *If X,Y are independent, then $Corr(X,Y) = 0$.*

Remark 3.53. The converse is not true.
X a $N(0,1)$ RV and $Y = X^2$. We have

$$Cov(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = \mathbb{E}[X^3] - \mathbb{E}[X]\mathbb{E}[X^2] = 0$$

3.3 Conditional Distribution

Recall that if $\mathbb{P}(B) > 0$ then

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

The conditional distribution of X given by Y is to study the distribution of X based on the information on Y.

Definition 3.54. X,Y are RVs on $(\Omega, \mathcal{S}, \mathbb{P})$. The conditional DF of X given $Y = y$ is defined as

$$F_{X|Y}(x|y) = \lim_{\epsilon \downarrow 0} \mathbb{P}(X \leq x : Y \in (y-\epsilon, y+\epsilon)) = \lim_{\epsilon \downarrow 0} \frac{\mathbb{P}(X \leq x, Y \in (y-\epsilon, y+\epsilon))}{\mathbb{P}(Y \in (y-\epsilon, y+\epsilon))}$$

if the limit exists.

3.3.1 Case 1

(X, Y) is a continuous RV with PDF $f_{(X,Y)}(x, y)$. Assume that the marginal PDFs are f_X, f_Y.

$$\frac{\mathbb{P}(X \leq x, Y \in (y-\epsilon, y+\epsilon))}{\mathbb{P}(Y \in (y-\epsilon, y+\epsilon))} = \frac{\int_{-\infty}^{x} \int_{y-\epsilon}^{y+\epsilon} f_{(X,Y)}(u, v) dv du}{\int_{y-\epsilon}^{y+\epsilon} f_Y(v) dv} = \frac{\int_{-\infty}^{x} \frac{\int_{y-\epsilon}^{y+\epsilon} f_{(X,Y)}(u,v) dv du}{2\epsilon}}{\frac{\int_{y-\epsilon}^{y+\epsilon} f_Y(v) dv}{2\epsilon}}$$

$$\xrightarrow{\epsilon \downarrow 0} \frac{\int_{-\infty}^{x} f_{(X,Y)}(u, y) du}{f_Y(y)}$$

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

Define $f_{X|Y}(x|y) = \frac{f_{(X,Y)}(x,y)}{f_Y(y)}$, then $F_{X|Y}(x|y) = \int_{-\infty}^{x} f_{X|Y}(u|y) du$.
$f_{X|Y}(x|y)$ is the conditional PDF of X given $Y = y$.

Example 3.55.

$$\mathbb{P}(X > 1 | Y = y) = \int_{1}^{\infty} f_{X|Y}(u|y) du; \quad \mathbb{P}(X \in A | Y = y) = \int_{A} f_{X|Y}(u|y) du$$

Definition 3.56. If $\forall y \in \mathbb{R}, \int_{\mathbb{R}} |x| f_{X|Y}(x|y) dx < \infty$ then the conditional expectation of X given Y denoted $\mathbb{E}[X|Y]$ is defined to be the RV on Ω given by:
$\forall \omega \in \Omega$, if $Y(\omega) = y$, then

$$\mathbb{E}[X|Y](\omega) = \int_{\mathbb{R}} x f_{X|Y}(x|y) dx$$

Remark 3.57. $f_{X|Y}(x|y)$ is indeed a PDF for

$$\int_{\mathbb{R}} f_{X|Y}(x|y) dx = \frac{\int_{\mathbb{R}} f_{(X,Y)}(x, y) dx}{f_Y(y)} = \frac{f_Y(y)}{f_Y(y)} = 1$$

Proposition 3.58.

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

Proof. Define $\int x f_{X|Y}(x, Y) dx = g(Y)$

$$\mathbb{E}[\mathbb{E}[X|Y]] = \mathbb{E}\left[\int_{\mathbb{R}} x f_{X|Y}(x, Y) dy\right] = \mathbb{E}[g(Y)]$$

$$= \int_{\mathbb{R}} g(y) f_Y(y) dy = \int_{\mathbb{R}} \int_{\mathbb{R}} x f_{X|Y}(x|y) dx f_Y(y) dy = \int_{\mathbb{R}} x \int_{\mathbb{R}} f_{(X,Y)}(x, y) dy dx$$

$$= \int_{\mathbb{R}} f_X(x) dx = \mathbb{E}[X]$$

$\square$

Proposition 3.59. *If X, Y are independent*

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

then

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

and

$$\forall \omega \in \Omega \quad \mathbb{E}[X|Y](\omega) = \int_{\mathbb{R}} x f_{X|Y}(x|Y(\omega))dx = \int_{\mathbb{R}} x f_X(x)dx = \mathbb{E}[X]$$

$\mathbb{E}[X|Y]$ *is a Dirac RV and* $\mathbb{E}[X|Y] = \mathbb{E}[X]$.

Example 3.60. $(X,Y) : \Omega \to \mathbb{R}^2$ continuous with joint PDF.

$$f_{(X,Y)}(x,y) = \frac{1}{2\pi} \exp\left(-\frac{2x^2 - 2xy + y^2}{2}\right)$$

First, determine f_X, f_Y.

$$f_X(x) = \int_{\mathbb{R}} \frac{1}{2\pi} \exp\left(-\frac{2x^2 - 2xy + y^2}{2}\right) dy = \frac{1}{2\pi} \int_{\mathbb{R}} \exp\left(-\frac{(y-x)^2}{2}\right) \exp\left(-\frac{x^2}{2}\right) dy$$

$$= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$$

This is the PDF of $N(0,1)$. Similarly, we have $f_Y(y) = \frac{1}{2\sqrt{\pi}} \exp\left(-\frac{y^2}{4}\right)$, the PDF of $N(0,2)$.

Second determine $f_{X|Y}$ and $\mathbb{E}[X|Y]$.

$$f_{X|Y}(x|y) = \frac{f_{(X,Y)}(x,y)}{f_Y(y)} = \frac{\frac{1}{2\pi} \exp\left(-\frac{2x^2-2xy+y^2}{2}\right)}{\frac{1}{2\sqrt{\pi}} \exp\left(-\frac{y^2}{4}\right)} = \frac{1}{\sqrt{\pi}} \exp\left(-\left(x - \frac{y}{2}\right)^2\right)$$

This is the PDF of $N\left(\frac{y}{2}, \frac{1}{2}\right)$.

$$\forall \omega \in \Omega, Y(\omega) = y \quad \mathbb{E}[X|Y](\omega) = \int_{\mathbb{R}} x f_{X|Y}(x|y)dx = \frac{y}{2} \implies \mathbb{E}[X|Y](\omega) = \frac{Y(\omega)}{2}$$

$$\mathbb{E}[\mathbb{E}[X|Y]] = \mathbb{E}\left[\frac{Y}{2}\right] = 0 = \mathbb{E}[X]$$

Example 3.61. Determine $\mathbb{E}[X|Y]$ for $(X,Y):\Omega \to (0,\infty)^2$

$$f_{(X,Y)}(x,y) = \begin{cases} \frac{e^{-\frac{x}{y}}e^{-y}}{y}: & \text{if } x,y>0 \\ 0: & \text{otherwise} \end{cases} \tag{3.20}$$

$$f_Y(y) = \int_{\mathbb{R}} f_{(X,Y)}(x,y)dx = \int_0^\infty e^{-\frac{x}{y}}dx\frac{e^{-y}}{y} = e^{-y}$$

if $y>0$

$$\implies f_Y(y) = \begin{cases} e^{-y}: & \text{if } y>0 \\ 0: & \text{otherwise} \end{cases} \tag{3.21}$$

$$f_{X|Y}(x|y) = \begin{cases} \frac{1}{y}e^{-\frac{1}{y}x}: & \text{if } x,y>0 \\ 0: & \text{if } x\le 0, y>0 \\ undefined: & \text{otherwise} \end{cases} \tag{3.22}$$

$\forall \omega \in \Omega$, if $Y(\omega)=y$.

$$\mathbb{E}[X|Y](\omega) = \int_{\mathbb{R}} x f_{X|Y}(x|y)dx = \int_0^\infty x\frac{1}{y}e^{-\frac{x}{y}}dx$$

Also, note that the conditional distribution of X given $Y=y, y>0$ is an exponential distribution with parameter $\frac{1}{y}$.
Then, we know $\mathbb{E}[X|Y](\omega) = y \implies \mathbb{E}[X|Y]=Y$.
What is $\mathbb{P}(X>1|Y=y)$?

$$\mathbb{P}(X>1|Y=y) = \int_1^\infty f_{X|Y}(x|y)dx = 1-F(1) = e^{-\frac{1}{y}}$$

3.3.2 Case 2

(X,Y) is discrete, $X\in\{x_i : i\in\mathbb{N}\}, Y\in\{y_j : j\in\mathbb{N}\}$.

$$F_{X|Y}(x|y) = \lim_{\epsilon\downarrow 0}\frac{\mathbb{P}(X\le x, Y\in(y-\epsilon,y+\epsilon))}{\mathbb{P}(Y\in(y-\epsilon,y+\epsilon))}$$

Remark 3.62. Since $Y\in\{y_j : j\in\mathbb{N}\}$, if $y=y_j$ then $\{y\in(y-\epsilon,y+\epsilon)\}\xrightarrow{\epsilon\downarrow 0}\{y_j\}$ and if $y\notin\{y_j : j\in\mathbb{N}\}$ then $\{y\in(y-\epsilon,y+\epsilon)\}\xrightarrow{\epsilon\downarrow 0}\emptyset$.
So $F_{X|Y}(x|y)$ is only defined if $y=y_j$ for some $j\in\mathbb{N}$ and

$$F_{X|Y}(x|y_j) = \frac{\mathbb{P}(X\le x, Y=y_j)}{\mathbb{P}(Y=y_j)} = \mathbb{P}(X\le x|Y=y_j)$$

Since $\mathbb{P}(X \leq x, Y = y_j) = \sum_{\{i \in \mathbb{N}: x_i \leq x\}} \mathbb{P}(X = x_i, Y = y_j)$

$$F_{X|Y}(x|y_j) = \sum_{\{i \in \mathbb{N}: x_i \leq x\}} \mathbb{P}(X = x_i | Y = y_j)$$

This is the conditional PMF of X given $Y = y_j$.

Definition 3.63. If $\sum_{i=1}^{\infty} |x_i| \mathbb{P}(X = x_i | Y = y_j) < \infty$, $\forall y_j$ then $\mathbb{E}[X|Y]$ is defined to be the RV on Ω such that $\forall \omega \in \Omega$, if $Y(\omega) = y_j$ then

$$\mathbb{E}[X|Y](\omega) = \sum_{i=1}^{\infty} x_i \mathbb{P}(X = x_i | Y = y_j)$$

In general, if $g : \mathbb{R} \to \mathbb{R}$ is Borel then

$$\mathbb{E}[g(X)|Y](\omega) = \sum_{i=1}^{\infty} g(x_i) \mathbb{P}(X = x_i | Y = y_j)$$

Proposition 3.64.

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

Proof. $\mathbb{E}[X|Y]$ is also a discrete RV.

$$\mathbb{E}[X|Y] \in \left\{ \sum_{i=1}^{\infty} x_i \mathbb{P}(X = x_i | Y = y_j) : j \in \mathbb{N} \right\}$$

Denote $\sum_{i=1}^{\infty} x_i \mathbb{P}(X = x_i | Y = y_j) = e_j$

$$\mathbb{P}(\mathbb{E}[X|Y] = e_j) = \mathbb{P}(Y = y_j)$$

This is the PMF of $\mathbb{E}[X|Y]$

$$\mathbb{E}[\mathbb{E}[X|Y]] = \sum_{j=1}^{\infty} e_j \mathbb{P}(\mathbb{E}[X|Y] = e_j)$$

$$= \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} x_i \mathbb{P}(X = x_i | Y = y_j) \mathbb{P}(Y = y_j)$$

$$= \sum_{i=1}^{\infty} x_i \sum_{j=1}^{\infty} \mathbb{P}(X = x_i, Y = y_j) = \mathbb{E}[X]$$

□

Remark 3.65. If X, Y are independent then

$$\mathbb{P}(X = x_i | Y = y_j) = \frac{\mathbb{P}(X = x_i)\mathbb{P}(Y = y_j)}{\mathbb{P}(Y = y_j)} = \mathbb{P}(X = x_i)$$

$$\implies \mathbb{E}[X|Y] = \mathbb{E}[X]; \quad \forall \omega \in \Omega$$

Proposition 3.66. *Given any Borel $g : \mathbb{R} \to \mathbb{R}$ if $X = g(Y)$ then $\mathbb{E}[g(Y)|Y] = g(Y)$.*

Proof. (Proof of the Y discrete case)

$X \in \{x_k := g(y_k) : k \in \mathbb{N}\}$ The joint PMF $\mathbb{P}(X = x_k, Y = y_j) = 0$ if $x_k \neq g(y_j)$. If $Y(\omega) = y_j$ then

$$\begin{aligned}
\mathbb{E}[X|Y](\omega) &= \sum_{k=1}^{\infty} x_k \mathbb{P}(X = x_k | Y = y_j) \\
&= \sum_{k=1}^{\infty} x_k \frac{\mathbb{P}(X = x_k, Y = y_j)}{\mathbb{P}(Y = y_j)} \\
&= \sum_{\{k : x_k = g(y_j)\}} x_k \frac{\mathbb{P}(X = x_k, Y = y_j)}{\mathbb{P}(Y = y_j)} \\
&= g(y_j) \sum_{\{k : x_k = g(y_j)\}} \frac{\mathbb{P}(X = x_k, Y = y_j)}{\mathbb{P}(Y = y_j)} \\
&= g(y_j) \\
\implies \mathbb{E}[g(Y)|Y] &= g(Y)
\end{aligned}$$

□

Example 3.67. Assume that within a day the number of trucks carrying gravel arriving at a factory is a $P(\lambda)$ RV, and the weight of the gravel carried by the trucks are IID $N(m, \sigma^2)$ RVs.
Assume the weight of the gravel carried by each truck is independent of the number of trucks arriving at the factory.
What is the expectation of the total weight of the gravel that arrives in one day?

Proof. N is a $P(\lambda)$ RV, $\{X_i : i \in \mathbb{N}\}$ are IID $N(m, \sigma^2)$ RVs. $\{X_i\}, N$ are independent.
$S_n = \sum_{i=1}^{n} X_i, S_N : \omega \in \Omega \mapsto S_N(\omega) = \sum_{i=1}^{n} X_i(\omega)$ is $N(\omega) = n$.

$$\mathbb{E}[S_N] = \mathbb{E}[\mathbb{E}[S_N | N]]$$

We do this, as it is easier to compute. The conditional DF is

$$F_{S_N|N}(x|n) = \frac{\mathbb{P}(S_N \le x, N = n)}{\mathbb{P}(N = n)} = \frac{\mathbb{P}(S_n \le x, N = n)}{\mathbb{P}(N = n)}$$

$$= \frac{\mathbb{P}(S_n \le x)\mathbb{P}(N = n)}{\mathbb{P}(N = n)} = \mathbb{P}(S_n \le x) = F_{S_n}(x)$$

S_n is Gaussian and is a $N(mn, \sigma^2 n)$ RV.
The conditional distribution of S_N given $N = n$ is $N(mn, \sigma^2 n)$. If $N(\omega) = n$, $\mathbb{E}[S_N|N](\omega) = nm \implies \mathbb{E}[S_N|N](\omega) = Nm$.
So we have

$$\mathbb{E}[S_N] = \mathbb{E}[\mathbb{E}[S_N|N]]\mathbb{E}[Nm] = m\mathbb{E}[N] = \lambda m$$

$\square$

Remark *3.68.* More generally, given $\{X_i : i \in \mathbb{N}\}$ IID RVs with $\mathbb{E}[X_i] = m$ and another N RV, $N \in \mathbb{N}$ and N independent of $\{X_i\}$ and $\mathbb{E}[N] < \infty$ then $\mathbb{E}[S_N] = \mathbb{E}[X_i]\mathbb{E}[N]$.

4 Asymptotics/The Law of Large Numbers

This is the study of the behaviour of a large number of RVs. (”limit”)

4.1 Modes of Convergence

Throughout 4.1 we will have $\{X_n : n \in \mathbb{N}\}$ and X are RVs on $(\Omega, \mathcal{S}, \mathbb{P})$, the DF of X_n is $F_n, n \ge 1$, the DF of X is F.

Definition 4.1. If for every $x \in \mathbb{R}$ that is a continuous point of F – i.e. F is continuous at x – we have that $\lim_{n\to\infty} F_n(x) = F(x)$, then we say that X_n converges to X in distribution (or in law) and write $X_n \to X$ in distribution. Equivalently, we say that $F_n \to F$ weakly. This is the weakest type of convergence.

Remark *4.2.* In general, F_n does not have to converge; if F_n converges then it does not have to converge to a distribution function.

Example *4.3.*

$$F_n(x) = \begin{cases} 0 : & \text{if } x < n \\ 1 : & \text{if } x \ge n \end{cases} \tag{4.1}$$

We have

$$\forall x \in \mathbb{R} \quad \lim_{n\to\infty} F_n(x) = 0$$

This is not a DF.

Example 4.4.

$$F_n(x) = \begin{cases} 0: & \text{if } x < \frac{1}{n} \\ 1: & \text{if } x \geq \frac{1}{n} \end{cases} \tag{4.2}$$

We have

$$\lim_{n\to\infty} F_n(x) = F(x) = \begin{cases} 0: & \text{if } x \leq 0 \\ 1: & \text{if } x > 0 \end{cases} \tag{4.3}$$

F is continuous everywhere except 0, so on $\mathbb{R}^\times$ we have $\lim_{n\to\infty} F_n(x) = F(x)$, so $X_n \to \delta_0$ in distribution.

Remark 4.5. Convergence in distribution only depends on the distribution of X_n and X. If $X_n \to X$ in distribution and X and Y have the same distribution and X, Y have the same distribution then $X_n \to Y$ in distribution.

Example 4.6. X is a $N(0,1)$ RV. Set $X_n = X, \forall n \geq 1$. Then $X_n \to X, -X$ in distribution.

Remark 4.7. $X_n \to X$ in distribution $\not\Longrightarrow$ $f_n \to f$ where f_n and f are the respective distributions.

Example 4.8. $X_n = \delta_{\frac{1}{n}}$ We have

$$\mathbb{P}(X_n = x) = \begin{cases} 1: & \text{if } x = \frac{1}{n} \\ 0: & \text{otherwise} \end{cases} ; \quad \mathbb{P}(X = x) = \begin{cases} 1: & \text{if } x = 0 \\ 0: & \text{otherwise} \end{cases} \tag{4.4}$$

We then have

$$\forall x \in \mathbb{R}, \quad \lim_{n\to\infty} \mathbb{P}(X_n = x) = 0 \neq \mathbb{P}(X = x)$$

Theorem 4.9. *If $X_n \in \mathbb{N}, \forall n \geq 1$ and $X \in \mathbb{N}$ then $X_n \to X$ in distribution $\iff$ $\forall k = 0, 1, 2, ...$ we have $\lim_{n\to\infty} \mathbb{P}(X_n = k) = \mathbb{P}(X = k)$.*

Proof. ($\Longrightarrow$) $\forall k = 0, 1, 2, ...$ we have

$$\mathbb{P}(X_n = k) = \mathbb{P}\left(X_n \in \left(k - \frac{1}{2}, k + \frac{1}{2}\right]\right)$$

$$= F_n\left(k + \frac{1}{2}\right) - F_n\left(k - \frac{1}{2}\right) \xrightarrow{n\to\infty} F\left(k + \frac{1}{2}\right) - F\left(k - \frac{1}{2}\right) = \mathbb{P}(X = k)$$

($\Longleftarrow$) $\mathbb{R} - \mathbb{N}$ are all the continuous points of F.

$$\forall x \in \mathbb{R} - \mathbb{N}, x \in (k-1, k) \quad F_n(x) = \sum_{l=0}^{k-1} \mathbb{P}(X_n = l) \xrightarrow{n\to\infty} \sum_{l=0}^{k-1} \mathbb{P}(X = l) = F(x)$$

$\square$

Theorem 4.10. *Let X_n, X be continuous RVs with PDFs f_n, f. Then if* $\lim_{n\to\infty} f_n(x) = f(x)$ *for almost every $x \in \mathbb{R}$ then $X_n \to X$ in distribution.*

Proof. $\forall x \in \mathbb{R}$

$$|F_n(x) - F(x)| = \left|\int_{-\infty}^{x} (f_n(t) - f(t))dt\right| \leq \int_{-\infty}^{x} |f_n(t) - f(t)|dt \to 0$$

$\square$

Remark 4.11. $X_n \to X$ in distribution $\not\Longrightarrow$ $f_n \to f$.

$$F_n(x) = \begin{cases} 0: & \text{if } x \leq 0 \\ x - \frac{\sin(2n\pi x)}{2n\pi}: & \text{if } 0 < x < 1 \\ 1: & \text{if } x \geq 1 \end{cases} \tag{4.5}$$

We have

$$\lim_{n\to\infty} F_n(x) = F(x) = \begin{cases} 0: & \text{if } x \leq 0 \\ x: & \text{if } 0 < x < 1 \\ 1: & \text{if } x \geq 1 \end{cases} \tag{4.6}$$

This is the DF of $U([0,1])$ with

$$f(x) = \begin{cases} 0: & \text{if } x \leq 0, x \geq 1 \\ 1: & \text{if } x \in (0,1) \end{cases} \tag{4.7}$$

but $f_n \not\to f$.

Proposition 4.12. *If $X_n \to X$ in distribution; $a, b \in \mathbb{R}$ then $aX_n + b \to aX + b$ in distribution.*

Proof. Exercise. $\square$

Definition 4.13. If $\forall \epsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}(|X_n - X| > \epsilon) = 0$$

then we say that $X_n \to X$ in probability.

Example 4.14. $\mathbb{P}(X_n = 1) = \frac{1}{n}, \mathbb{P}(X_n = 0) = 1 - \frac{1}{n}$ then $\forall \epsilon > 0$

$$\mathbb{P}(|X_n - 0|) > \epsilon = \mathbb{P}(X_n = 1) = \frac{1}{n} \xrightarrow{n\to\infty} 0$$

So X_n goes to 0 in probability.

Remark 4.15. $X_n \to X$ in probability $\iff$ $X_n - X \to 0$ in probability.

Proposition 4.16. *If $X_n \to X$ in probability and $X_n \to Y$ in probability then $\mathbb{P}(X = Y) = 1$.*

Proof.
$$\{|X - Y| > \epsilon\} \subset \left\{|X_n - X| > \frac{\epsilon}{2}\right\} \cup \left\{|X_n - Y| > \frac{\epsilon}{2}\right\}$$
This is because $\epsilon < |X - Y| \leq |X - X_n| + |Y - Y_n|$ So now $\forall \epsilon > 0$

$$\mathbb{P}\left(|X - Y| > \epsilon\right) \leq \mathbb{P}\left(\left\{|X_n - X| > \frac{\epsilon}{2}\right\} \cup \left\{|X_n - Y| > \frac{\epsilon}{2}\right\}\right)$$
$$\leq \mathbb{P}\left(|X_n - X| > \frac{\epsilon}{2}\right) + \mathbb{P}\left(|X_n - Y| > \frac{\epsilon}{2}\right) \xrightarrow{n\to\infty} 0$$
$$\implies \mathbb{P}(|X - Y| > \epsilon) = 0$$
$$\mathbb{P}(X \neq Y) = \mathbb{P}\left(\bigcup_{k=1}^{\infty} \left\{|X - Y| > \frac{1}{k}\right\}\right) \leq \sum_{k=1}^{\infty} \mathbb{P}\left(|X - Y| > \frac{1}{k}\right)$$
□

Proposition 4.17. *If $X_n \to X, Y_n \to Y$ in probability then $X_n + Y_n; X_n - Y_n; Y_n - X_n; X_n \cdot Y_n \to X + Y; X - Y; Y - X; XY$ in probability.*

Proof. $\forall \epsilon > 0$
$$\mathbb{P}(|X_n + Y_n - (X + Y)| > \epsilon) \leq \mathbb{P}\left(|X_n - X| > \frac{\epsilon}{2}\right) + \mathbb{P}\left(|Y_n - Y| > \frac{\epsilon}{2}\right) \xrightarrow{n\to\infty} 0$$
□

Remark 4.18. This does not hold for convergence in distribution.
$X_n = X, Y_n = -X$ then we have $X_n \to X, Y_n \to X$ in distribution but $0 = X_n + Y_n \neq 2X$.

Theorem 4.19. *$X_n \to X$ in probability $\implies$ $X_n \to X$ in distribution.*

Proof. Let F_n, F be the DFs of X_n, X. $\forall x \in \mathbb{R}$
$$\{X_n \leq x\} \subset \{X_n \leq x, X \leq x + \epsilon\} \cup \{X_n \leq x, X > x + \epsilon\}$$
$$\subset \{X \leq x + \epsilon\} \cup \{|X_n - X| > \epsilon\}$$
First we have
$$\mathbb{P}(X_n \leq x) \leq \mathbb{P}(X \leq x + \epsilon) + \mathbb{P}(|X_n - X| > \epsilon)$$
Repeating the argument we get
$$\mathbb{P}(X \leq x + \epsilon) \leq \mathbb{P}(X_n \leq x) + \mathbb{P}(|X_n - X| > \epsilon)$$

Combining the two we get

$$F(x-\epsilon) - \mathbb{P}(|X_n - X| > \epsilon) \leq F_n(x) \leq F(x+\epsilon) + \mathbb{P}(|X_n - X| > \epsilon)$$

Then we get

$$F(x-\epsilon) - F(x) - \mathbb{P}(|X_n - X| > \epsilon) \leq F_n(x) - F(x) \leq F(x+\epsilon) - F(x) + \mathbb{P}(|X_n - X| > \epsilon)$$

If x is a continuous point of F then $\forall \epsilon' > 0$ by choosing ϵ small enough one can make $|F(x) - F(y)| < \epsilon'$ if $|x - y| < \epsilon \implies F(x-\epsilon) - F(x) > -\epsilon'; F(x+\epsilon) - F(x) < \epsilon'$.
With the chosen $\epsilon > 0$, since $X_n \to X$, by choosing n large enough we can make $-\epsilon' < \mathbb{P}(|X_n - X| > \epsilon) < \epsilon'$. This then implies

$$|F_n(x) - F(x)| \leq 2\epsilon'$$

So $\lim_{n\to\infty} F_n(x) = F(x) \implies X_n \to X$ in distribution. □

Remark 4.20. Warning! The opposite is not true in general, e.g.,
X is $N(0,1), X_n = -X, \forall n \geq 1$. $X_n \to X$ in distribution, but $-X = X_n \not\to X$ in prob.

Theorem 4.21. *(Slutsky's Theorem)*
If $X_n \to X$ in distribution, $Y_n \to c$ in prob for some $c \in \mathbb{R}$ then $X_n + Y_n \to X + C$ in distribution and $X_n Y_n \to cX$ in distribution.

Proof. The first statement is an exercise.
For the second statement we will first show that $X_n(Y_n - c) \to 0$ in prob. $\forall \epsilon > 0, \forall M > 0$

$$\mathbb{P}(|X_n(Y_n - c)| > \epsilon) = \mathbb{P}\left(|X_n(Y_n - c)| > \epsilon, |Y_n - c| \leq \frac{\epsilon}{M}\right) + \mathbb{P}\left(|X_n(Y_n - c)| > \epsilon, |Y_n - c| > \frac{\epsilon}{M}\right)$$

$$\leq \mathbb{P}(|X_n| > M) + \mathbb{P}\left(|Y_n - c| > \frac{\epsilon}{M}\right)$$

$$\leq |\mathbb{P}(|X_n| > M) - \mathbb{P}(|X| > M)| + \mathbb{P}(|X| > M) + \mathbb{P}\left(|Y_n - c| > \frac{\epsilon}{M}\right)$$

Now $\forall \epsilon' > 0$ Now for

$$\mathbb{P}(|X| > M)$$

Since $X \in \mathbb{R}$ one can find large enough M such that $\mathbb{P}(|X| > M) < \epsilon'$. For

$$\mathbb{P}\left(|Y_n - c| > \frac{\epsilon}{M}\right)$$

Since $Y_n \to c$ in probability, for our chosen M we can choose n large enough that $\mathbb{P}\left(|Y_n - c| > \frac{\epsilon}{M}\right) < \epsilon'$. Finally for

$$|\mathbb{P}(|X_n| > M) - \mathbb{P}(|X| > M)|$$

we first assume WLOG that $M, -M$ are continuous points of F. Then we have

$$\mathbb{P}(|X_n| > M) - \mathbb{P}(|X| > M) = -\mathbb{P}(|X_n| \leq M) + \mathbb{P}(|X| \leq M)$$
$$= F(M) - F_n(M) - (F(-M) - F_n(-M)) \xrightarrow{n\to\infty} 0$$

Thus we have

$$\leq |\mathbb{P}(|X_n| > M) - \mathbb{P}(|X| > M)| + \mathbb{P}(|X| > M) + \mathbb{P}\left(|Y_n - c| > \frac{\epsilon}{M}\right) \xrightarrow{n\to\infty} 0$$

So we have $X_nY_n - cX_n \to 0$ in prob and thus $X_nY_n = X_nY_n - cX_n + cX_n \to cX$ in distribution by the first statement. □

Definition 4.22. FYI (not tested)

$X_n \to X$ almost surely (a.s.) if $\mathbb{P}(\lim_{n\to\infty} X_n = X) = 1$.
$X_n \to X$ in L^p ($p \geq 1$) if $\lim_{n\to\infty} \mathbb{E}[|X_n - X|^p] = 0$.

Remark 4.23. In general, there is no relationship between convergence a.s. and convergence in L^p but we have both
$X_n \to X$ a.s. $\implies X_n \to X$ in prob $\implies X_n \to X$ in dist.
and
$X_n \to X$ in $L^p \implies X_n \to X$ in prob $\implies X_n \to X$ in dist.
Convergence in L^p implies convergence in prob because $\forall \epsilon > 0$

$$\mathbb{P}(|X_n - X| > \epsilon) = \mathbb{P}(|X_n - X|^p > \epsilon^p) \leq \frac{\mathbb{E}[|X_n - X|^p]}{\epsilon^p} \to 0$$

Note that this is Markov's inequality.

4.2 Weak Law of Large Numbers

Remark 4.24. For $\{X_i : i \in \mathbb{N}\}$ each a RV on $(\Omega, \mathcal{S}, \mathbb{P})$ then we define $\forall n \geq 1$

$$S_n = \sum_{i=1}^n X_i; \quad \overline{S_n} = \frac{S_n}{n}$$

Definition 4.25. We say that $\{X_i\}$ obeys the WLLN if $\frac{S_n - \mathbb{E}[S_n]}{n} \to 0$ in prob $\iff$ $\overline{S_n} - \mathbb{E}[\overline{S_n}] \to 0$ in prob.

Theorem 4.26. *WLLN(1)*
(recall HW4 Q7)
If $\{X_i : i \in \mathbb{N}\}$ is uncorrelated ($Cov(X_i, X_j) = 0, \forall i \neq j$) and the second moments are bounded ($\sup_{i\in\mathbb{N}} \mathbb{E}[X_i^2] = c < \infty$) then $\{X_i\}$ has the WLLN.

Lemma 4.27. *If $\{X_i : i \in \mathbb{N}\}$ is uncorrelated and $\mathbb{E}[X_i^2] < \infty$ then $Var(S_n) = \sum_{i=1}^n Var(X_i)$.*

Proof.

$$\begin{aligned}
Var(S_n) = \mathbb{E}[(S_n - \mathbb{E}[S_n])^2] &= \mathbb{E}\left[\left(\sum_{i=1}^n (X_i - \mathbb{E}[X_i])\right)^2\right] \\
&= \mathbb{E}\left[\sum_{i=1}^n \sum_{j=1}^n (X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])\right] \\
= \sum_{i=1}^n \mathbb{E}[(X_i - \mathbb{E}[X_i])^2] &+ \sum_{i \neq j} \mathbb{E}[(X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])] \\
= \sum_{i=1}^n Var(X_i) &+ \sum_{k \neq j} Cov(X_k, X_j) = \sum_{i=1}^n Var(X_i)
\end{aligned}$$

□

Proof. $\forall \epsilon > 0$

$$\mathbb{P}\left(\left|\frac{S_n - \mathbb{E}[S_n]}{n}\right| > \epsilon\right) = \mathbb{P}\left(|S_n - \mathbb{E}[S_n]|^2 > n^2\epsilon^2\right)$$

by Chebychev

$$\begin{aligned}
&\leq \frac{\mathbb{E}[|S_n - \mathbb{E}[S_n]|^2]}{n^2\epsilon^2} = \frac{Var(S_n)}{n^2\epsilon^2} = \frac{\sum_{i=1}^n Var(X_i)}{n^2\epsilon^2} \\
&\leq \frac{cn}{n^2\epsilon^2} \to 0
\end{aligned}$$

□

Corollary 4.28. *If $\{X_i\}$ are uncorrelated IIDs with $\mathbb{E}[X_i] = m, Var(X_i) = \sigma^2$ then $\overline{S_n} - \mathbb{E}[\overline{S_n}] \to 0$ in prob., i.e. $\overline{S}_n \to m$ in prob.*

Example 4.29. $\{X_i\}$ are IID $U([0,1])$ then $\overline{S_n} \to \frac{1}{2}$ in prob.

Remark 4.30. Variance depends on the 2nd moment. Study the proof to understand how to deal with this problem.

Theorem 4.31. *WLLN(2)*
If $\{X_i : i \in \mathbb{N}\}$ are pairwise independent ($\implies$ uncorrelated) and identically distributed with the first moment being finite, i.e. $\mathbb{E}[X_i] = m$ then $\{X_i\}$ has WLLN, i.e. $\overline{S_n} \to m$ in prob.

Proof. Fix any $A > 0$, define $Y_i = \chi_{\{|X_i| \le A\}} X_i$. Then we have

$$|Y_i| \le A \implies Var(Y_i) \le \mathbb{E}[Y_i^2] \le A^2$$

$\{Y_i\}$ is also pairwise independent-identically distributed, so WLLN applie to $\{Y_i\}$, i.e. $\frac{\sum_{i=1}^n Y_i}{n} \to m_A$ in prob., where $m_A = \mathbb{E}[Y_i] = \mathbb{E}[X_i \chi_{\{|X_i| \le A\}}]$. Now $\forall \epsilon > 0, \forall 0 < \epsilon' < \frac{\epsilon}{2}$

$$|\overline{S_n} - m| \le \left|\overline{S_n} - \frac{\sum_{i=1}^n Y_i}{n}\right| + \left|\frac{\sum_{i=1}^n Y_i}{n} - m_A\right| + |m_A - m|$$

For the third term we have

$$\mathbb{E}[X_i \chi_{\{|X_i| \le A\}}] \le \mathbb{E}[|X_i| \chi_{\{|X_i| \le A\}}]$$

because $\mathbb{E}[|X_i|] < \infty$ we have

$$\lim_{A \to \infty} \mathbb{E}[|X_i| \chi_{\{|X_i| > A\}}] = \lim_{A \to \infty} (\mathbb{E}[|X_i|] - \mathbb{E}[|X_i| \chi_{\{|X_i| \le A\}}]) = 0$$

We can choose A large enough so that $|m_A - m| \le \mathbb{E}[|X_i| \chi_{\{|X_i| > A\}}] \le \epsilon'$. Then

$$\begin{aligned} \mathbb{P}(|\overline{S_n} - m| > \epsilon) &\le \mathbb{P}\left(\left|\overline{S_n} - \frac{\sum_{i=1}^n Y_i}{n}\right| + \left|\frac{\sum_{i=1}^n Y_i}{n} - m_A\right| > \frac{\epsilon}{2}\right) \\ &\le \mathbb{P}\left(\left|\overline{S_n} - \frac{\sum_{i=1}^n Y_i}{n}\right| > \frac{\epsilon}{4}\right) + \mathbb{P}\left(\left|\frac{\sum_{i=1}^n Y_i}{n} - m_A\right| > \frac{\epsilon}{4}\right) \end{aligned}$$

For the second term we have, with A chosen, $\frac{\sum_{i=1}^n Y_i}{n} - m_A \to 0$ in prob. $\implies$ one can choose n large enough so that

$$\mathbb{P}\left(\left|\frac{\sum_{i=1}^n Y_i}{n} - m_A\right| > \frac{\epsilon}{4}\right) \le \epsilon'$$

Finally, for the first term we have

$$\overline{S_n} - \frac{\sum_{i=1}^n Y_i}{n} = \frac{\sum_{i=1}^n (X_i - Y_i)}{n} = \frac{\sum_{i=1}^n X_i \chi_{|X_i| > A}}{n}$$

which implies, also with Markov's inequality

$$\begin{aligned} \mathbb{P}\left(\left|\overline{S_n} - \frac{\sum_{i=1}^n Y_i}{n}\right| > \frac{\epsilon}{4}\right) &= \mathbb{P}\left(\frac{\sum_{i=1}^n X_i \chi_{|X_i| > A}}{n} > \frac{\epsilon}{4}\right) \le \frac{\mathbb{E}\left[\sum_{i=1}^n X_i \chi_{|X_i| > A}\right]}{\frac{\epsilon}{4} n} \\ &\le \frac{\sum_{i=1}^n \mathbb{E}[X_i \chi_{|X_i| > A}]}{\frac{\epsilon}{4} n} = \frac{n \mathbb{E}[X_i \chi_{|X_i| > A}]}{\frac{\epsilon}{4} n} \le \frac{\epsilon' \epsilon}{\frac{\epsilon}{4}} = 4\epsilon' \end{aligned}$$

This implies that

$$\mathbb{P}(|\overline{S_n} - m| > \epsilon) < 5\epsilon'$$

when n is large $\implies \overline{S_n} \to m$ in prob. $\square$

Remark 4.32. WLLN does not always hold, e.g. X_i is an $N(0,i)$ RV, $\{X_i\}$ independent then $\overline{S_n}$ is a $N\left(0, \frac{\sum_{i=1}^n i}{n^2}\right)$ RV. $Var(\overline{S_n}) = \frac{n(n+1)}{2n^2} \approx \frac{1}{2}$, so $\overline{S_n} \approx N(0, \frac{1}{2})$ when n is large, thus $\overline{S_n} \not\to 0$ in prob.

4.3 Strong Law of Large Numbers

Definition 4.33. $\{X_i : i \in \mathbb{N}\}$ are uncorrelated with the fourth moment bounded, (i.e. $\sup_{i\in\mathbb{N}} \mathbb{E}[X_i^4] = c < \infty$) then $\frac{S_n - \mathbb{E}[S_n]}{n} \to 0$ a.s., i.e., $\mathbb{P}(\lim_{n\to\infty} \frac{S_n-\mathbb{E}[S_n]}{n} = 0) = 1$.

$\{X_i : i \in \mathbb{N}\}$ are IID with $\mathbb{E}[X_i] = m \in \mathbb{R}$ then $\overline{S_n} \to m$ a.s. (i.e. $\mathbb{P}(\lim_{n\to\infty} \overline{S_n} = m) = 1$).

Theorem 4.34. *(Weierstrass Approximation Theorem)*

Let f be continuous on $[0,1]$. $\forall n \geq 1$ define $p_n(x) = \sum_{k=0}^n f\left(\frac{k}{n}\right)\binom{n}{k} x^k (1-x)^{n-k}$ (Bernstein's polynomial).

$$\forall x \in [0,1], \lim_{n\to\infty} p_n(x) = f(x) \quad \textit{i.e.} \lim_{n\to\infty} |p_n(x) - f(x)| = 0$$

$$\lim_{n\to\infty} \sup_{x\in[0,1]} |p_n(x) - f(x)| = 0 \quad \textit{i.e. } p_n \textit{ converges to } f \textit{ uniformly on } [0,1]$$

Proof. For the first statement, $\forall x \in [0,1]$ fixed consider $\{X_i : i \in \mathbb{N}\}$ IID Bernoulli $B(x)$ RVs, i.e. $\mathbb{P}(X_i = 1) = x, \mathbb{P}(X_i = 0) = 1 - x$. WLLN $\implies \frac{S_n}{n} \to x$ in prob. From HW5 we have

$$f\left(\frac{S_n}{n}\right) \to f(x) \text{ in prob.} \quad \implies \mathbb{E}\left[f\left(\frac{S_n}{n}\right)\right] \to f(x)$$

Recall that S_n is $B(n,x)$ so

$$\mathbb{E}\left[f\left(\frac{S_n}{n}\right)\right] = \sum_{k=0}^n f\left(\frac{k}{n}\right)\binom{n}{k} x^k(1-x)^{n-k} = p_n(x) \implies \lim_{n\to\infty} p_n(x) = f(x)$$

□

Proof. For the second statement we have

$$|p_n(x) - f(x)| = \left|\mathbb{E}\left[f\left(\frac{S_n}{n}\right) - f(x)\right]\right| \leq \mathbb{E}\left[\left|f\left(\frac{S_n}{n}\right) - f(x)\right|\right]$$

Because f is bounded and uniformly continuous on $[0,1]$, we have $\sup_{x\in[0,1]} |f(x)| = A < \infty$. $\forall \epsilon > 0, \exists \delta > 0$ such that $|f(x) - f(y)| \leq \epsilon$ for every $x, y \in [0,1]$ with $|x - y| \leq \delta$. Now continuing the estimate we have

$$\mathbb{E}\left[\left|f\left(\frac{S_n}{n}\right) - f(x)\right|\right] = \mathbb{E}\left[\left|f\left(\frac{S_n}{n}\right) - f(x)\right|\chi_{\{|\frac{S_n}{n} - x| \leq \delta\}}\right] + \mathbb{E}\left[\left|f\left(\frac{S_n}{n}\right) - f(x)\right|\chi_{\{|\frac{S_n}{n} - x| > \delta\}}\right]$$

$$\leq \epsilon + 2A\mathbb{P}\left(\left|\frac{s_n}{n} - x\right| > \delta\right) \leq \epsilon + 2A\frac{Var\left(\frac{S_n}{n}\right)}{\delta^2}$$

$$= \epsilon\frac{2A}{n^2\delta^2}Var(S_n) \leq \epsilon + \frac{2A}{n^2\delta^2}nVar(X_i) = \epsilon + \frac{2A}{n\delta^2}x(1-x)$$

Thus we get

$$\sup_{x\in[0,1]} |p_n(x) - f(x)| \leq \epsilon + \frac{2A}{n\delta^2}\sup_{x\in[0,1]} x(1-x)$$

$$= \epsilon + \frac{A}{2n\delta^2} \xrightarrow{n\to\infty} \epsilon$$

$$\implies \lim_{n\to\infty}\sup_{x\in[0,1]}|p_n(x) - f(x)| = 0$$

□

4.4 Central Limit Theorem

Remark 4.35. Study the role of MGFs in convergence in distribution.

Example 4.36. Recall that if X_n is a Dirac RV at $\frac{1}{n}$, i.e. $\mathbb{P}\left(X_n = \frac{1}{n}\right) = 1$ we know that $X_n \to X$ in distribution where X is δ_0. $\forall s \in \mathbb{R}$

$$M_n(s) := M_{X_n}(s) = e^{s\frac{1}{n}} \xrightarrow{n\to\infty} 1$$

The MGF of X, $M(s) \equiv 1$. Thus we have

$$M_n(s) \to M(s)$$

Example 4.37. $\{X_i : i \in \mathbb{N}\}$ IID RVs with $\mathbb{E}[X_i] = m \in \mathbb{R}$. Assume that the MGF $M_i(s)$ of X_i exists for $s \in (-s_0, s_0)$. Then we have

$$M_{\overline{S_n}}(s) = \mathbb{E}\left[e^{s\frac{\sum_{i=1}^n X_i}{n}}\right] = \prod_{i=1}^n \mathbb{E}[e^{\frac{s}{n}X_i}] = \prod_{i=1}^n M_i\left(\frac{s}{n}\right) = \left(M_i\left(\frac{s}{n}\right)\right)^n$$

Taylor expand at M_1 to get

$$M_i\left(\frac{s}{n}\right) = 1 + M_i'(0)\frac{s}{n} + o\left(\frac{1}{n}\right)$$

$$\implies M_{\overline{S_n}}s = \left(1 + m\frac{s}{n} + o\left(1n\right)\right)^n \xrightarrow{n\to\infty} e^{ms} = M(s)$$

$M(s)$ is the MGF of δ_m.
Recall that WLLN implies that $\overline{S_n} \to m$ in prob.

Theorem 4.38. *(Levy's Continuity Theorem)*

Let $\{X_n\}$ and X be RVs with the MGFs M_n, M. If $\exists s_0 > 0$ such that $\forall s \in (-s_0, s_0)$; $M_n(s), M(s)$ exist and $\lim_{n\to\infty} M_n(s) = M(s)$, then $X_n \to X$ in dist. (convergence of MGF in a neighbourhood of 0 $\implies$ convergence in distribution)

Remark 4.39. In general $X_n \to X$ in distribution does not mean that $M_n \to M$. For examples, assume X_n has DF

$$F_n(x) = \begin{cases} 0: & \text{if } x < -n \\ \frac{1}{2} + c_n \arctan(nx): & \text{if } x \in [-n, n] \quad \text{where } c_n = \frac{1}{2\arctan(n^2)} \xrightarrow{n\to\infty} \frac{1}{\pi} \\ 1: & \text{otherwise} \end{cases} \tag{4.8}$$

We have then

$$\lim_{n\to\infty} F_n(x) = \begin{cases} 0: & \text{if } x < 0 \\ \frac{1}{2}: & \text{if } x = 0 \\ 1: & \text{otherwise} \end{cases} \tag{4.9}$$

Notice that if X is δ_0

$$F(x) = \begin{cases} 0: & \text{if } x < 0 \\ 1: & \text{otherwise} \end{cases} \tag{4.10}$$

At every continuous point of F we have $\lim_{n\to\infty} F_n(x) = F(x) \implies X_n \to X$ in distribution. We have the PDF

$$f_n(x) = \begin{cases} 0: & \text{if } x < -n \\ \frac{c_n n}{1+n^2x^2}: & \text{if } x \in [-n, n] \\ 0: & \text{if } x > n \end{cases} \tag{4.11}$$

and so we have

$$M_n(s) = \int_{-n}^{n} e^{sx} \frac{c_n n}{1+n^2x^2} dx = \int_{-n^2}^{n^2} e^{\frac{sy}{n}} \frac{c_n}{1+y^2} dy \geq \int_{\frac{n^2}{2}}^{n^2} e^{\frac{sy}{n}} \frac{dy}{1+y^2} c_n$$

$$\geq e^{\frac{sn}{2}} c_n \frac{\frac{n^2}{2}}{1+n^2} \geq C e^{\frac{sn}{2}} \xrightarrow[\forall s>0]{n\to\infty} \infty$$

Thus $M_n \not\to M$.

Example 4.40. Given $\lambda > 0, \forall n \in \mathbb{N}$ let $\{X_i : i = 1, ..., n\}$ IID Bernoulli $B\left(\frac{\lambda}{n}\right)$ RVs. Then $\forall s \in \mathbb{R}$

$$M_{S_n}(s) = \mathbb{E}\left[e^{s\sum_{i=1}^n X_i}\right] = \prod_{i=1}^{n} \mathbb{E}[e^{sX_i}] = (M_{X_i}(s))^n = \left(e^s\frac{\lambda}{n} + 1 - \frac{\lambda}{n}\right)^n$$

$$= \left(1 + (e^s - 1)\frac{\lambda}{n}\right)^n \xrightarrow{n\to\infty} e^{\lambda(e^s-1)} = M(s)$$

This is the MGF of $P(\lambda)$. Thus $S_n \to Y$ in dist. where Y is $P(\lambda)$.

Example 4.41. Let $\{X_i : i \in \mathbb{N}\}$ be IID $B\left(\frac{1}{2}\right)$. Define $\forall n \in \mathbb{N}$

$$Z_N := \sum_{i=1}^{n} \frac{X_i}{2^i} \in [0,1]$$

We claim $Z_n \to Y$ where Y is $U([0,1])$.

Proof.

$$M_{Z_n}(s) = \prod_{i=1}^{n} M_{X_i}\left(\frac{s}{2^i}\right) = \prod_{i=1}^{n}\left(\frac{e^{\frac{s}{2^i}}+1}{2}\right)$$

By induction we have, $\forall n \geq 1$

$$\prod_{i=1}^{n}\left(\frac{e^{\frac{s}{2^i}}+1}{2}\right) = (e^s - 1)\frac{1}{2^n\left(e^{\frac{s}{2^n}}-1\right)}$$

Also we have

$$\lim_{n\to\infty} \frac{\frac{1}{2^n}}{e^{\frac{s}{2^n}}-1} = \frac{1}{s}$$

$$\implies M_{Z_n}(s) \to \frac{e^s-1}{s}$$

This is the MGF of $U([0,1])$. ∎

Remark 4.42. Variance is how much randomness there is.

4.4.1 Set Up for the Central Limit Theorem(CLT)

$\{X_i : i \in \mathbb{N}\}$ independent with

$$\mathbb{E}[X_i] = m_i \in \mathbb{R}; \quad Var(X_i) = \sigma_i^2 < \infty; \quad \tau_n^2 = \sum_{i=1}^{n}\sigma_i^2 \implies Var(S_n) = \tau_n^2$$

Consider

$$T_n = \frac{S_n - \mathbb{E}[S_n]}{\tau_n}$$

where $\tau_n = \sqrt{\tau_n^2}$. We have

- $\forall n \geq 1, \mathbb{E}[T_n] = 0$ and $Var(T_n) = \frac{Var(S_n)}{\tau_n^2} = 1$ i.e. there is a fixed amount of randomness.

- If $\{X_i : i \in \mathbb{N}\}$ are IID $N(m, \sigma^2)$ then S_n is $N(nm, n\sigma^2)$ and T_n is $N(0,1)$.
- CLT describes the phenomenon that – under certain conditions – even if $X_i's$ are not Gaussian, T_n resembles an $N(0,1)$ RV where n is large, i.e. $T_n \to Z$ in dist. where Z is $N(0,1)$.

4.4.2 Characterizations

Theorem 4.43. *CLT(1)*

Let $\{X_i : i \in \mathbb{N}\}$ be a sequence of IID RVs with $\mathbb{E}[X_i] = m, Var(X_i) = \sigma^2$. Then $\{X_i\}$ obeys the CLT, i.e. $T_n \to Z$ in dist. where Z is $N(0,1)$. In this case

$$T_n = \frac{S_n - \mathbb{E}[S_n]}{\tau_n} = \frac{S_n - nm}{\sqrt{n}\sigma} \to Z \text{ in dist.}$$

In particular, $\forall x \in \mathbb{R}$

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{S_n - nm}{\sqrt{n}\sigma} \le x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt = \Phi(x)$$

More generally, $\forall a < b$

$$\lim_{n\to\infty} \mathbb{P}\left(a < \frac{S_n - nm}{\sqrt{n}\sigma} \le b\right) = \frac{1}{\sqrt{2\pi}} \int_{a}^{b} e^{-\frac{t^2}{2}} dt = \Phi(b) - \Phi(a)$$

Proof. Assume that the MGF $M_i = M_{X_i}$ exists in a neighbourhood of 0.

$$M_{T_n}(s) = \mathbb{E}\left[e^{s\frac{S_n - nm}{\sqrt{n}\sigma}}\right] = e^{-\frac{sm\sqrt{n}}{\sigma}} \mathbb{E}\left[e^{\frac{sS_n}{\sqrt{n}\sigma}}\right] = e^{-\frac{sm\sqrt{n}}{\sigma}} \prod_{i=1}^{n} M_{X_i}\left(\frac{s}{\sqrt{n}\sigma}\right)$$

$$= e^{-\frac{sm\sqrt{n}}{\sigma}} \left(M_{X_i}\left(\frac{s}{\sqrt{n}\sigma}\right)\right)^n = \left(e^{-\frac{sm}{\sqrt{n}\sigma}} M_{X_i}\left(\frac{s}{\sqrt{n}\sigma}\right)\right)^n$$

We Taylor expand

$$e^{-\frac{sm}{\sqrt{n}\sigma}} = 1 - \frac{sm}{\sqrt{n}\sigma} + \frac{s^2m^2}{2n\sigma^2} + o\left(\frac{1}{n}\right)$$

$$M_{X_i}\left(\frac{s}{\sqrt{n}\sigma}\right) = 1 + m\frac{s}{\sqrt{n}\sigma} + \frac{(\sigma^2 + m^2)}{2}\frac{s^2}{n\sigma^2} + o\left(\frac{1}{n}\right)$$

Taking the product, we get

$$e^{-\frac{sm}{\sqrt{n}\sigma}} M_{X_i}\left(\frac{s}{\sqrt{n}\sigma}\right) = 1 - \frac{sm}{\sqrt{n}\sigma} + \frac{s^2m^2}{2n\sigma^2} + m\frac{s}{\sqrt{n}\sigma} + \frac{(\sigma^2 + m^2)}{2}\frac{s^2}{n\sigma^2} - \frac{s^2m^2}{n\sigma^2} + o\left(\frac{1}{n}\right)$$

So

$$M_{T_n}(s) = \left(1 + \frac{s^2}{2n} + o\left(\frac{1}{n}\right)\right)^n \to e^{\frac{s^2}{2}}$$

This is the MGF of $N(0,1)$. $\square$

Example 4.44. $\{X_i : i \in \mathbb{N}\}$ IID $B\left(\frac{1}{2}\right)$ RVs. Estimate $\mathbb{P}(S_n = 20)$ where $n = 40$. S_n is $B\left(n, \frac{1}{2}\right)$.

$$\mathbb{P}(S_n = 20) = \binom{n}{20}\left(\frac{1}{2}\right)^n = \binom{40}{20}\left(\frac{1}{2}\right)^{40} \approx 0.1254$$

$$T_n = \frac{S_n - \frac{n}{2}}{\sqrt{n}\frac{1}{2}} \to Z \text{ in dist.}$$

Where Z is $N(0,1)$.

$$\mathbb{P}(S_n = 20) = \mathbb{P}(19.5 < S_n < 20.5) = \mathbb{P}\left(\frac{19.5 - \frac{n}{2}}{\sqrt{n}\frac{1}{2}} < T_n < \frac{20.5 - \frac{n}{2}}{\sqrt{n}\frac{1}{2}}\right) = \mathbb{P}\left(\frac{-1}{2\sqrt{10}} < T_{40} < \frac{1}{2\sqrt{10}}\right)$$

$$\approx \Phi\left(\frac{1}{2\sqrt{10}}\right) - \Phi\left(\frac{-1}{2\sqrt{10}}\right) \approx 0.1272$$

Example 4.45. $\{X_i : i \in \mathbb{N}\}$ IID $B\left(\frac{1}{2}\right)$ RVs. Estimate $\mathbb{P}\left(S_n \leq \frac{n}{2}\right)$ where n is large.

$$\mathbb{P}\left(S_n \leq \frac{n}{2}\right) = \mathbb{P}(T_n \leq 0) = \sum_{\left\{k \in \mathbb{N}, k \leq \frac{n}{2}\right\}} \binom{n}{k}\frac{1}{2^n} \to \Phi(0) = \frac{1}{2}$$

Remark 4.46. $\{X_i : i \in \mathbb{N}\}$ IID with $\mathbb{E}[X_i] = m, Var(X_i) = \sigma^2$. Then $\forall \epsilon > 0$

$$\mathbb{P}(|\overline{S_n} - m| > \epsilon) = \mathbb{P}\left(\frac{|S_n - nm|}{\sqrt{n}\sigma} > \frac{\sqrt{n}\epsilon}{\sigma}\right) = \mathbb{P}\left(|T_n| > \frac{\sqrt{n}\epsilon}{\sigma}\right)$$

Now by CLT, when n is large

$$\mathbb{P}\left(|T_n| > \frac{\sqrt{n}\epsilon}{\sigma}\right) \approx \mathbb{P}\left(|Z| > \frac{\sqrt{n}\epsilon}{\sigma}\right)$$

where Z is $N(0,1)$.

$$\mathbb{P}\left(|Z| > \frac{\sqrt{n}\epsilon}{\sigma}\right) = \mathbb{P}\left(e^{cZ^2} > e^{\frac{cn\epsilon^2}{\sigma^2}}\right) \leq \frac{\mathbb{E}\left[e^{cZ^2}\right]}{e^{\frac{cn\epsilon^2}{\sigma^2}}} = \frac{1}{\sqrt{1-2c}}e^{n\frac{-c\epsilon^2}{\sigma^2}}$$

Thus we have

$$\mathbb{P}(|\overline{S_n} - m| > \epsilon) = \mathbb{P}\left(|T_n| > \frac{\sqrt{n}\epsilon}{\sigma}\right) \leq \frac{1}{\sqrt{1-2c}}e^{n\frac{-c\epsilon^2}{\sigma^2}}$$

This is exponential decay, so we have $\overline{S_n} \to m$ in prob.

Remark 4.47. For IID RVs CLT $\implies$ WLLN. CLT also tells you how fast $\mathbb{P}(|\overline{S_n} - m| > \epsilon)$ decays, exp. fast.

Theorem 4.48. *CLT(2)*

Let $\{X_i : i \in \mathbb{N}\}$ *be a sequence of IID RVs with*

$$\mathbb{E}[X_i] = m_i; \quad Var(X_i) = \sigma_i^2; \quad \tau_n^2 = \sum_{i=1}^{n} \sigma_i^2 = Var(S_n)$$

If $\forall \epsilon > 0$ *we have the Lindenberg Condition*

$$\frac{1}{\tau_n^2} \sum_{i=1}^{n} \mathbb{E}\left[|X_i - m_i|^2 \chi_{\{|X_i - m_i| > \epsilon \tau_n\}}\right] \xrightarrow{n\to\infty} 0$$

then $T_n = \frac{S_n - \mathbb{E}[S_n]}{\tau_n} \to N(0,1)$ *in dist.*

Remark 4.49. If $\{X_i : i \in \{1,...,n\}\}$ are IID RVs and $\mathbb{E}[X_i] = m, Var(X_i) = \sigma^2 \implies \tau_n^2 = n\sigma^2$ then $\forall \epsilon > 0$ the Lindenberg condition gives

$$\frac{1}{n\sigma^2} \sum_{i=1}^{n} \mathbb{E}\left[|X_i - m|^2 \chi_{\{|X_i - m| > \epsilon\sqrt{n}\sigma\}}\right] = \frac{1}{\sigma^2}\mathbb{E}\left[|X_i - m|^2 \chi_{\{|X_i - m| > \epsilon\sqrt{n}\sigma\}}\right] \xrightarrow{n\to\infty} 0$$

because

$$\mathbb{E}\left[|X_i - m|^2 \chi_{\{|X_i - m| > \epsilon\sqrt{n}\sigma\}}\right] = \mathbb{E}\left[|X_i - m|^2\right] - \mathbb{E}\left[|X_i - m|^2 \chi_{\{|X_i - m| \le \epsilon\sqrt{n}\sigma\}}\right]$$

and the second part of the equality converges to 0. Thus we have that CLT(2) $\implies$ CLT(1).

Example 4.50. Let $\{X_i : i \in \mathbb{N}\}$ be a sequence of independent RVs and $|X_i| \le A, \forall i \in \mathbb{N}$ where A is a fixed constant, i.e. the X_i's are uniformly bounded. Now

$$\sup_{i\in\mathbb{N}} \mathbb{E}[X_i^2] \le A^2$$

So $\{X_i : i \in \mathbb{N}\}$ have bounded 2nd moments, thus by WLLN we have $\frac{S_n - \mathbb{E}[S_n]}{n} \to 0$ in prob.

If $\tau_n^2 = Var(S_n) = \sum_{i=1}^{n} Var(X_i) \to \infty$ then we have $\forall \epsilon > 0$

$$\frac{1}{\tau_n^2} \sum_{i=1}^{n} \mathbb{E}\left[|X_i - m_i|^2 \chi_{\{|X_i - m_i| > \epsilon \tau_n\}}\right] \le \frac{4A^2}{\tau_n^2} \sum_{i=1}^{n} \mathbb{P}(|X_i - m_i| > \epsilon \tau_n) \le \frac{4A^2}{\tau_n^2} \sum_{i=1}^{n} \frac{Var(X_i)}{\epsilon^2 \tau_n^2}$$

$$= \frac{4A^2}{\epsilon^2 \tau_n^2} \to 0$$

Thus we have CLT.

If $\tau_n^2 \to \tau_2 < \infty$ then Lindenberg fails. We can find $\epsilon > 0$ such that $\mathbb{P}(|X_i - m_i| > \epsilon\tau_n) = \delta > 0$ and so

$$\frac{1}{\tau_n^2}\sum_{i=1}^{n}\mathbb{E}\left[|X_i - m_i|^2\chi_{\{|X_i-m_i|>\epsilon\tau_n\}}\right] \geq \frac{1}{\tau_n^2}\mathbb{E}\left[|X_i - m_i|^2\chi_{\{|X_i-m_i|>\epsilon\tau_n\}}\right]$$

$$\geq \frac{\epsilon^2\tau_n^2}{\tau_n^2}\mathbb{P}(|X_i - m_i| > \epsilon\tau_n) \geq \epsilon^2\mathbb{P}(|X_i - m_i| > \epsilon\tau) = \epsilon^2\delta > 0$$

Thus the Lindenberg condition fails and so CLT may fail.

Example 4.51. This is an example of CLT failing. Define X_i s.t.

$$\mathbb{P}\left(X_i = \frac{1}{i^2}\right) = \frac{1}{2} = \mathbb{P}\left(X_i = -\frac{1}{i^2}\right)$$

with X_i independent, $Var(X_i) = \frac{1}{i^4}, \mathbb{E}[X_i] = 0$. Thus we have $\sum_{i=1}^{\infty}\frac{1}{i^4} < \infty$

$$T_n = \frac{S_n - 0}{\tau_n} = \frac{\sum_{i=1}^{n} X_i}{\sum_{i=1}^{n}\frac{1}{i^4}} \implies |T_n| \leq \sum_{i=1}^{\infty}\frac{1}{i^2} = c < \infty \implies \mathbb{P}(T_n > c) = 0 \not\to 1 - \Phi(c)$$

Remark 4.52. It's possible that CLT applies but WLLN doesn't.

Example 4.53. $\{X_i : i \in \mathbb{N}\}$ independent with $\mathbb{P}(X_i = i^\lambda) = \frac{1}{2} = \mathbb{P}(X_i = -i^\lambda)$ with $\lambda > 0$. We have $\mathbb{E}[X_i] = 0, Var(X_i) = i^{2\lambda}$ and

$$\tau_n^2 = \sum_{i=1}^{n} Var(X_i) = \sum_{i=1}^{n} i^{2\lambda} \approx \int_1^n x^{2\lambda}dx = o\left(n^{2\lambda+1}\right)$$

Fix $\epsilon > 0$. One can find large N such that $\forall n \geq N, \forall i = 1, ..., n$ we have $i^\lambda \leq n^\lambda \leq \epsilon\tau_n$. Since $\tau_n = o\left(n^{\lambda+\frac{1}{2}}\right)$ we have

$$|X_i| \leq \epsilon\tau_n \implies \{|X_i| > \epsilon\tau_n\} = \emptyset$$

Now

$$\frac{1}{\tau_n^2}\sum_{i=1}^{n}\mathbb{E}\left[|X_i - m_i|^2\chi_{\{|X_i-m_i|>\epsilon\tau_n\}}\right] = \frac{1}{\tau_n^2}\sum_{i=1}^{N-1}\mathbb{E}\left[|X_i - m_i|^2\chi_{\{|X_i-m_i|>\epsilon\tau_n\}}\right]$$

$$\leq \frac{(N-1)(N-1)^{2\lambda}}{\tau_n^2} \xrightarrow{n\to\infty} 0$$

Thus we have $\forall a \in \mathbb{R}$

$$\mathbb{P}\left(a < \frac{S_n}{\tau_n}\right) \xrightarrow{n\to\infty} 1 - \Phi(a)$$

$$\implies \mathbb{P}\left(\frac{a\tau_n}{n} < \frac{S_n}{n}\right) \xrightarrow{n\to\infty} 1 - \Phi(a)$$

Choose a such that $\Phi(a) < 1$. Now if $\lambda \geq \frac{1}{2}$, for M constant and large enough n we have

$$\frac{\tau_n}{n} = o\left(n^{\lambda+\frac{1}{2}}\right)\frac{1}{n} \geq M$$

so for large n

$$\mathbb{P}\left(aM < \frac{S_n}{n}\right) \geq 1 - \Phi(a)$$

$\implies \frac{S_n}{n} \not\to 0$ in prob. Thus WLLN does not hold.

Example 4.54. $\{X_i : i \in \mathbb{N}\}$ IID with $\mathbb{E}[X_i] = m, Var(X_i) = \sigma^2$ and $\mathbb{E}[X_i^2] < \infty$. Find the limiting distribution of

$$Z_n = \sqrt{n}\frac{X_1X_2 + X_3X_4 + \cdots + X_{2n-1}X_{2n} - nm^2}{\sum_{i=1}^{2n} X_i^2}$$

Solution:

$$Z_n = \sqrt{n}\frac{1}{\sum_{i=1}^{2n} X_i^2}(X_1X_2 + X_3X_4 + \cdots + X_{2n-1}X_{2n} - nm^2)$$

(1) $\{X_i^2\}$ IID, $\mathbb{E}[X_i^2] = \sigma^2 + m^2 \implies$ WLLN(2) $\implies \frac{\sum_{i=1}^{2n} X_i^2}{2n} \to \sigma^2 + m^2$ in prob. $\implies \frac{1}{\sum_{i=1}^{2n} X_i^2} \to \frac{2n}{\sum_{i=1}^{2n} X_i^2}$.

(2) $\{X_{2i-1}X_{2i} : i \in \mathbb{N}\}$ are IID because if for $i \neq j$, $\{X_{2i-1}, X_{2i}, X_{2j-1}, X_{2j}\}$ are independent then $X_{2i-1}X_{2i}$ is independent of $X_{2j-1}X_{2j}$.

$$\mathbb{E}[X_{2i-1}X_{2i}] = \mathbb{E}[X_{2i-1}]\mathbb{E}[X_{2i}] = m^2$$

$$\mathbb{E}[X_{2i-1}^2X_{2i}^2] = (m^2+\sigma^2)^2 \implies Var(X_{2i-1}X_{2i}) = \sigma^4 + 2m^2\sigma^2$$

Now by CLT

$$\frac{X_1X_2 + X_3X_4 + \cdots + X_{2n-1}X_{2n}}{\sqrt{n}\sqrt{\sigma^4 + 2m^2\sigma^2}} \to Z \text{ in dist.}$$

Where Z is $N(0,1)$.

Now by Slutsky's Thm.

$$Z_n \to \frac{1}{\sigma^2 + m^2} \frac{\sqrt{\sigma^4 + 2m^2\sigma^2}}{2} Z \text{ in dist.}$$

Thus the limiting distribution of Z is $N\left(0, \frac{\sigma^4+2m^2\sigma^2}{4(\sigma^2+m^2)^2}\right)$.