Probability Theory

Aamod Varma

MATH - 3235, Fall 2025

Contents

Chapter 1

Introduction

Example. What is the probability that two people among N people have the same birthday. $\diamond$

Example. What is the probability that all people have different birthday
We have,

$$\begin{aligned}
q_1 &= 1 \\
q_2 &= \left(1 - \frac{1}{365}\right) \\
q_3 &= q_3\left(1 - \frac{2}{365}\right) \\
&\vdots \\
q_n &= \prod_{i=1}^{n-1}\left(1 - \frac{i}{365}\right)
\end{aligned}$$

We get $q_n = 0.14$ which gives us 0.86 for the previous example.

$\diamond$

Note. We assume certain assumptions like the following to make this work,

1. Uniformity
2. Independence

Here we have a probability model and deduced the probability of an event,

Example. Say there is a test for a disease,

1. P(positive | sick) = 1
2. P(positive | not sick) = 0.01

Need to find P(sick | positive) which would be P(positive | sick) P(sick) / P(positive)
We test everybody, we have Assume 100 S and 100 NS,
100 P from the S, 99 P from the NS
So we have 199 P of which only 100 S which gives around .5

$\diamond$

1.1 Probability Theory

Experiment whose outcome is not determined. We define the following,

1. Ω : Sample space, set of possible outcomes

Example. (a) Throw a die,
$\Omega = \{1, 2, 3, 4, 5, 6\} \to$ finite

(b) Flip a coin till heads,
$\Omega = \{1, 2, 3, \dots\} = \mathbb{N} \to$ countably infinite

(c) Time to wait till next bus arrival,
$\Omega = \mathbb{R}^+ \to$ uncountabaly infinite

$\diamond$

2. F : Family of events, $A, B, \dots$
   Something that may or may not happen
   Example. (a) For a die we can ask,
   - Is the outcome even?
   - Is the outcome ≤ 3?

   Here an event $A \subseteq \Omega$ and $|\Omega| = 6$ so $|2^\Omega| = 64$
   We have $F =$ family of events $= 2^\Omega$

   (b) Here we have,
   $\Omega = \mathbb{N}$ so $F = 2^\mathbb{N}$

   (c) In this case our sample space is $R^+ = (0, \infty)$. But we cannot take $2^\mathbb{R}$. So we axiomatically define F as noted below. Under this definition F is the smallest family that contains all open intervals of R

   $\diamond$

3. P : How likely an event is

Definition (Axiomatic definition of F)**.** So here we define F to be a family of events of Ω if,

1. not empty
2. if $A \in F \Rightarrow A^c \in F$ $(A^c = \Omega \setminus A)$
3. for any two $A, B \in F$ then $A \cup B \in F$
4. If A_i for $i = 1, \dots, \infty$ are events, then $\bigcup_{i=1}^{\infty} A_i$ is an event

Note. Here, countable closure $\Rightarrow$ finite closure (proof just involves adding infinite ϕ to our finite sets $A_1, \dots, A_n$)

Note. Using this definition we have,

1. $A \in F \Rightarrow A^c \in F, \Rightarrow A \cup A^c = \Omega \in F$ and $\phi = \Omega^c \in F$
   So every event space has Ω, ϕ
2. $(A \cup B)^c = A^c \cap B^c \in F$ so,
   If $A_i, i = 1, 2, \dots$ are events then we have,
   $(\bigcap_{i=1}^{\infty} A_i)^c \in F = \bigcup_{i=1}^{\infty} A_i^c \in F$

1.2 Probability

Definition (Axiomatic definition of Probability). A probability is a function $\mathbb{P} : F \to [0, 1]$ with the following probabilities, We want the following properties,

1. $\mathbb{P}(A) \geq 0$
2. $\mathbb{P}(\Omega) = 1$ and $\mathbb{P}(\phi) = 0$
3. If A & B are events, they are mutually exclusive if $A \cup B = \phi$ so it should have,

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

If A_i for $i = 1, 2, 3, \ldots$ are events with $A_i \cap A_j$ where $i \neq j$ then,

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

Example. (a). For the die, we have $\mathbb{P}(\{i\})$ for $i \in \{1, \ldots, 6\}$. So if Ω is finite, then the probability is completely defined by $\mathbb{P}(\omega)$ for $\omega \in \Omega$, here $\{\omega\}$ is called in atomic event. If A is an event then we have,

$$\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(w)$$

In particular, $\mathbb{P}$ is called uniform if,

$$\mathbb{P}(\omega) = \frac{1}{|\Omega|}$$

(b). Coin flip.
We have our sample space as $\mathbb{N}$. First, let's say that $\mathbb{P}(H) = p$ and $\mathbb{P}(T) = q = 1 - p$. Let x be the number of flips to get first head and $x \in \mathbb{N}$.

$$\begin{aligned}
P(1) &= p \\
P(2) &= (1-p)p \\
&\cdots \\
P(n) &= (1-p)^{n-1}p
\end{aligned}$$

We have,

$$\begin{aligned}
\sum_{n=1}^{\infty} (1-p)^{n-1}p = p \sum_{m=0}^{\infty} (1-p)^m \\
&= p\frac{1}{1-(1-p)} = \frac{p}{p} \\
&= 1
\end{aligned}$$

Note. This is true, $\sum_{n=0}^{\infty} x^n = \frac{1}{1-x}$ if $|x| < 1$
So we have $\mathbb{P}(A) = \sum_{n \in A} \mathbb{P}(n)$
(c). Consider $[A, B] \subset R$, if we take, $(x, y) \subset [A, B]$ so we have,

$$\mathbb{P}([x, y]) = k(y - x)$$

and

$$\mathbb{P}([A, B]) = 1$$

this means that $k = \frac{1}{B-A}$ so,

$$\mathbb{P}([x, y]) = \frac{y - x}{B - A}$$

◇

Definition (Probability Space)**.** The probability space is defined by $(\Omega, \mathbb{F}, \mathbb{P})$ where Ω is a sample space, $\mathbb{F}$ is a family of events and $\mathbb{P}$ is a probability on $\mathbb{F}$

Some consequence are,
1. $\Omega = A \cup A^c$ and $A \cap A^c = \phi$. So,

$$\mathbb{P}(\Omega) = 1 = \mathbb{P}(A) + \mathbb{P}(A^c)$$

which gives us,

$$\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$$

2. As $\phi = \Omega^c \Rightarrow$ if $\mathbb{P}(\Omega) = 1 \Rightarrow \mathbb{P}(\phi) = 0$
3. Given A, B as events,

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

Proof. We know that $A = A \setminus B \cup (A \cap B)$ and $B = B \setminus A \cup (A \cap B)$

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

We can write,

$$A \cup B = (A \setminus B) \cup (B \setminus A) \cup (A \cap B)$$

This gives us,

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

So get,

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

□

1.3 Conditional Probability

Given A, B what is the probability of B if I know that A happened?

Theorem 1.1. Given B with $\mathbb{P}(B) > 0$ let $\mathbb{Q}(A) = \mathbb{P}(A|B)$. $\mathbb{Q}$ is a probability.

Proof. 1. $\mathbb{Q}(A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \geq 0$ so $\mathbb{Q}(A) \geq 0$
2. $\mathbb{Q}(\omega) = \frac{\mathbb{P}(\Omega \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B)}{\mathbb{P}(B)} = 1$
3.

$$\begin{aligned}\mathbb{Q}(\bigcup_{i=1}^{\infty} A_i) &= \frac{\mathbb{P}((\bigcup_{i=1}^{\infty} A_i) \cap B)}{\mathbb{P}(B)} \\ &= \frac{\mathbb{P}((\bigcup_{i=1}^{\infty} A_i \cap B))}{\mathbb{P}(B)} \\ &= \frac{\sum_{i=1}^{\infty} \mathbb{P}(A_i \cap B)}{\mathbb{P}(B)}\end{aligned}$$

□

$\mathbb{P}(A|B) = \mathbb{P}(A)$ then A is independent from B, this implies that, $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B) \Rightarrow \mathbb{P}(B|A) = \mathbb{P}(B)$

Definition. A and B are independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$

Note. This implies that $\mathbb{P}(A|B) = \mathbb{P}(A)$

Example. A and B are independent iff A and B^c are independent.
We can write $A = (A \cap B) \cup (A \cap B^c)$. So we have,

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

Now we can write,

$$\begin{aligned} \mathbb{P}(A \cap B^c) &= \mathbb{P}(A) - \mathbb{P}(A)\mathbb{P}(B) \\ &= \mathbb{P}(A)(1 - \mathbb{P}(B)) \\ &= \mathbb{P}(A)\mathbb{P}(B^c) \end{aligned}$$

$\diamond$

Consider if we have three events A, B, C. Then if we have,

$$\begin{aligned} \mathbb{P}(A \cap C) &= \mathbb{P}(A)\mathbb{P}(C) \\ \mathbb{P}(A \cap B) &= \mathbb{P}(A)\mathbb{P}(B) \\ \mathbb{P}(B \cap C) &= \mathbb{P}(B)\mathbb{P}(C) \end{aligned}$$

This is called mutually independent (not a good definition for independence)

Example. Let four possible outcomes be $\{1, 2, 3, 4\}$. Now if we have $A = \{1, 2\}, B = \{1, 3\}, C = \{2, 3\}$. This gives us,

$$\begin{aligned} \mathbb{P}(A \cap B) &= \frac{1}{4} \\ \mathbb{P}(A) = \mathbb{P}(B) &= \frac{1}{2} \end{aligned}$$

Now $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(\phi) = 0 \neq \mathbb{P}(A)\mathbb{P}(B \cap C)$
So if we want that $\mathbb{P}(A|B \cap C) = \mathbb{P}(A)$ then $\mathbb{P}(A \cap B \cap C) = \mathbb{P}(A)\mathbb{P}(B \cap C) = \mathbb{P}(A)\mathbb{P}(B)sw_{,0(}$a1) ! store value of c at the address in $a1\mathbb{P}(C)$

$\diamond$

Exercise. A, B, C are independent then $\mathbb{P}(A|B \cup C) = \mathbb{P}(A)$. We can write $B \cup C = (B \cap C^c) \cup (B \cap C) \cup (B^c \cap C)$

Proposition 1.2. In general, $A_i, i \in I$ of events. A_i are independent if $\forall J \subset I$ then,

$$\mathbb{P}(\bigcap_{j \in J} A_j) = \prod_{j \in J} \mathbb{P}(A_j)$$

Note. This implies that if $J_1, J_2 \subset I$ with $J_1 \cap J_2 = \phi$. Then any combination of $A_i, i \in J_1$ is independent to any combination of $A_i, i \in J_2$

Definition (Parition)**.** Assume a family of events A_i. We call it a partition if $\bigcup_i A_i = \Omega$ and $A_i \cap A_j = \phi, \forall i \neq j$.

Theorem 1.3. If B is an event and A_i is a partition, then

$$\mathbb{P}(B) = \sum_i \mathbb{P}(B|A_i)\mathbb{P}(A_i)$$

Proof. We write,

$$\begin{aligned} B &= \bigcup_i (B \cap A_i) \\ \mathbb{P}(B) &= \sum_i \mathbb{P}(B \cap A_i) \\ &= \sum_i \mathbb{P}(B|A_i)\mathbb{P}(A_i) \end{aligned}$$

□

Example. Consider two production lines,

1. 1000 items, 0.01 defective
2. 500 items, 0.02 defective

If all items are collected and pick one at random, what is the probability that that it is defective. If D is the event that the item is defective so we need to find $P(D)$ if we have I and II as both the production lines then we have,

$$\mathbb{P}(D) = \mathbb{P}(D|I)\mathbb{P}(I) + \mathbb{P}(D|II)\mathbb{P}(II) = 0.01 \times \frac{2}{3} + 0.02 \times \frac{1}{3} = \frac{0.04}{3}$$

We can also ask if an item is picked and it's defective, what is the probability that it is from line I. So we need to find $\mathbb{P}(I|D)$.

$$\begin{aligned} \mathbb{P}(I|D) = \frac{\mathbb{P}(I \cap D)}{\mathbb{P}(D)} &= \frac{\mathbb{P}(D|I)\mathbb{P}(I)}{\mathbb{P}(D)} \\ &= \frac{\mathbb{P}(D|I)\mathbb{P}(I)}{\mathbb{P}(D|I)\mathbb{P}(I) + \mathbb{P}(D|II)\mathbb{P}(II)} \\ &= \frac{0.01 \times \frac{2}{3}}{\frac{0.04}{3}} \\ &= \frac{1}{2} \end{aligned}$$

◇

Theorem 1.4 (Bayes Thoerem)**.** If A_i is a partition and B is an event. Then,

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\sum_j \mathbb{P}(B|A_j)\mathbb{P}(A_j)}$$

Proof.

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

And we have from the partition theorem that $\mathbb{P}(B) = \sum_j \mathbb{P}(B|A_j)\mathbb{P}(A_j)$. Plugging this back in gives us the theorem. $\square$

Given P_1, P_2 positive at the first and second test. Then what is $\mathbb{P}(P_1 \cap P_2|NS)$

1.4 Examples

Example. Coin flip arrival of the first H
Let A be the event that $n > 10$ and B be the event that $n > 13$ and C is the even that $n > 13$. We show that $\mathbb{P}(B|A) = \mathbb{P}(C)$
We compute $\mathbb{P}(A)$ first. So,

$$\begin{aligned}\mathbb{P}(A) = \sum_{n=11}^{\infty} \mathbb{P}(\{n\}) = \sum_{n=11}^{\infty} p(1-p)^{n-1} &= (1-p)^{10} \sum_{n=11}^{\infty} p(1-p)^{n-11} \\ &= (1-p)^{10}\end{aligned}$$

This is the probability that the first 10 flips are tails. Similarly $\mathbb{P}(B) = (1-p)^{13}$ and $\mathbb{P}(C) = (1-p)^3$
So $\mathbb{P}(B|A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}$
We have $B \subseteq A$ so it's the same as $\mathbb{P}(B)$ so we have,

$$\mathbb{P}(B|A) = \frac{\mathbb{P}(B)}{\mathbb{P}(A)} = \frac{(1-p)^{13}}{(1-p)^3} = (1-p)^{10} = \mathbb{P}(C)$$

$\diamond$

Example. Given a box with 9 balls and 3 are blue and 6 are red.
Let R_1 be the first draw being red so we have $\mathbb{P}(R_1) = 2/3$. If we don't reinsert the ball we have $\mathbb{P}(R_2|R_1) = 5/8$ and $\mathbb{P}(R_2|B_1) = 6/8$.
Suppose if you pick a blue ball you win 10 and a red will give you 0.

$$\mathbb{P}(R_2) = \frac{2}{3}\frac{5}{8} + \frac{6}{8}\frac{1}{3} = \frac{5}{12} + \frac{3}{12} = \frac{8}{12} = \frac{2}{3}$$

$\diamond$

Remark. The point here is before starting drawing, the probability of the first, second, etc draw of reds or blues are the same. The probability will change given information on what the previous draw is but in the beginning it shouldn't make a difference.

Example. Consider N flips of a fair coin. Sample space would be $\{H, T\}^N$. Here $H_i = \{\text{i'th flip is } H\}$. We have $|\Omega| = 2^N$. Some $\omega \in \Omega$ is a sequence $\omega = \{w_1, \ldots, w_N\}$. We can show that H_i is independent from H_j if $i \neq j$. We have,

$$\mathbb{P}(H_i) = \frac{2^{n-1}}{2^n} = \frac{1}{2}$$

Similarly

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

And $\mathbb{P}(H_i \cap H_j) = \frac{1}{4}$
So we see that their independent. So the collection $H_i, i = 1, 2, \ldots, N$ is an independent collection of events. $\diamond$

Remark. More general we have, given two subsets I_1 and $I_2 \subset \{1, \ldots, N\}$ such that $I_1 \cap I_2 = \phi$ then for events A and B are some specific outcomes for $i \in I_1$ and $i \in I_2$ respectively. Then both the events A, B are independent.
Here A, B are called cylinder sets.

Example. If you flip the coin infinitely many times we have, $= \{0,1\}^N$ which is uncountable. So we consider $I \subset N$ and take $\sigma_i \in \{0,1\}, i \in I$. So our cylinder set would be,

$$\{\omega : \omega_i = \sigma_i\}$$

so here $\omega = \{\omega_1, \omega_2, \dots, \omega_n, \dots\}$

$\diamond$

Example. What is the probability of the outcome of getting only zeroes - would be zero. But this isn't an event in our family of events. But we can consider an event $A_n = \{\text{The outcome of the first } n \text{ flips is } 0\}$. For a fair coin this is 2^{-n}. We can say that,

$$A = \bigcap_{n=1}^{\infty} A_n \qquad \text{as we have } A_{n+1} \subset A_n$$

$\diamond$

Theorem 1.5. If A_n is a decreasing collection s.t. $A_{n+1} \subset A_n$ and we have $A = \bigcap_{n=1}^{\infty}$ then we have,

$$\mathbb{P}(A) = \lim_{n\to\infty} \mathbb{P}(A)$$

Theorem 1.6. If A_n is increasing so $A_n \subset A_{n+1}$ and $A = \bigcup_{n=1}^{\infty} A_n$ then we have,

$$\mathbb{P}(A) = \lim_{n\to\infty} \mathbb{P}(A_n)$$

Proof. We have,

$$\begin{aligned}
A_1 \cup A_2 &= A_1 \cup (A_2 \setminus A_1) \\
A_1 \cup A_2 \cup A_3 &= A \cup (A_2 A_1) \cup (A_3 A_2) \\
\bigcup_{i=1}^{N} A_i &= A_1 \bigcup_{i=2}^{N} B_i \qquad B_i = A_i A_{i-1} \\
\mathbb{P}(\bigcup_{i=1}^{\infty} A_i) &= \mathbb{P}(A_1) + \sum_{i=2}^{\infty} \mathbb{P}(B_i) \\
&= \mathbb{P}(A_1) + \sum_{i=2}^{\infty} (\mathbb{P}(A_i) - \mathbb{P}(A_{i-1})) \\
&= \mathbb{P}(A_1) + \lim_{n\to\infty} \sum_{i=2}^{n} (\mathbb{P}(A_i) - \mathbb{P}(A_{i-1})) \\
&= \mathbb{P}(A_i) = \lim_{n\to\infty} (\mathbb{P}(A_n) - \mathbb{P}(A_1))
\end{aligned}$$

$\square$

Chapter 2

Random Vairables

2.1 Introduction

Definition. If we have $(\Omega, \mathscr{F}, \mathbb{P})$ then $X : \Omega \to \mathbb{R}$ is a discrete random variable if,

1. $X(\Omega)$ is finite or countable.
2. $\forall a \in \mathbb{R}$ $(\forall a \in X(\Omega))$
$$\{\omega : X(\omega) = a\} \in \mathscr{F}$$

Remark. Here Ω can be uncountable but the point is that the mapping in $\mathbb{R}, X(\Omega)$ has to be countable.

Remark. The point of the second condition is that for any value in R, there is some event in $\mathscr{F}$ which maps to that value. So this point guarantees measurability.

Example. Consider $Y =$ number of H in T tosses of a coin, possible values are clearly $\{0, 1, \ldots, N\}$. We have,

$$\begin{aligned}
\mathbb{P}(Y = 0) &= (1-p)^N \\
\mathbb{P}(Y = 1) &= Np(1-p)^{N-1} \\
\mathbb{P}(Y = 2) &= \binom{N}{2} p^2 (1-p)^{N-2} \\
\mathbb{P}(Y = y) &= \binom{N}{y} p^y (1-p)^{N-y}
\end{aligned}$$

◇

Remark. Here $p_Y(y) = \mathbb{P}(Y = y)$ is called the probability mass function of Y.

Definition. If X is a discrete random variable then if,

$$p(x) = \mathbb{P}(X = x)$$

Then $p(x)$ is called the probability mass function (pmf)

Remark. We also have,

$$\begin{aligned}
\sum_{x \in X(\Omega)} p(x) &= \sum_{x \in X(\Omega)} \mathbb{P}(X = x) \\
&= \mathbb{P}\Bigg(\bigcup_{x \in X(\Omega)} \{\omega \in \Omega : X(\omega) = x\} \Bigg) \\
&= \mathbb{P}(\Omega) = 1
\end{aligned}$$

2.2 Examples

Example. A random variable that takes only 2 values (conventionally represented with 0, 1) is called a **Bernoulli r.v**. The pmf of a Bernoulli r.v. is completely given by $p(1) = \mathbb{P}(X = 1)$ and $p(0) = 1 - p(1)$ $\diamond$

Example. If a random variable that takes values $0, \dots, N$ with p.m.f,

$$p(x) = \binom{N}{x} p^x (1-p)^{N-x}$$

is called a ***binomial r.v.*** This is a 2 parameter pmf (p, N).

$\diamond$

Remark. ***Binomial can be thought of as a sum of*** N ***independent Bernoulli r.v. with parameter*** p. In addition, it can be thought of as the number of successes in N independent Bernoulli trials with success probability p.

Example. I have a bowl with N blue balls and M red balls. If we extract n balls without reinsertion. Then, what is the probability of x blue.

Here the total possible outcomes is $\binom{M+N}{n}$. We want to select x blue balls and $n - x$ red balls. So possibility for blue is $\binom{N}{x}$ and for red is $\binom{M}{n-x}$. So,

$$p(x) = \frac{\binom{N}{x}\binom{M}{n-x}}{\binom{M+N}{n}}$$

This is called a **hypergeometric r.v.** which is a 3 parameter pmf. $\diamond$

Note. If $x \ll N$ then $\binom{N}{x} \sim N^x$

Remark. If n and x are fixed and take $N, M \to \infty$ then the hypergeometric pmf converges to a binomial.

Example. Geometric r.v.. Take $\Omega = \{\underline{\omega} = (\omega_1, \dots, \omega_n, \dots)\}$ where $\omega_i \in \{0, 1\}$

$X(\underline{\omega}) =$ the position of the first 1 in $\underline{\omega}$

$\{\underline{\omega} \mid X(\underline{\omega}) = n\} = A_n$
But A_n is the set of all $\underline{\omega}$ such that $\omega_1 = \omega_2 = \dots = \omega_{n-1} = 0$
Here $X(\Omega) = \mathbb{N}$ which means its countable and $X^{-1}(n) \in \mathscr{F}$ so it means it's a random variable.

$p(n) = \mathbb{P}(X = x) = p(1-p)^{n-1}$
If X has the p.m.f $p(x)$ above then it's called a **Geometric r.v.** with parameter p.

$\diamond$

Remark. ***Geometric can be thought of as the number of trials until the first success in a sequence of independent Bernoulli trials each with success probability*** p.

Remark. Here A_n is a cylinder set as we're fixing the value of ω on a finite number of points (in this case from $1, \dots, n-1$)

Remark. Sometimes q is taken as $1 - p$ so this is also correct, $p(x) = pq^{x-1}$

Example. Suppose you have a binomial with large N but $pN = \lambda$ is finite. So,

$$p(x) = \binom{N}{x} p^x (1-p)^{N-x}$$

From here we have $p = \frac{\lambda}{n}$ so we have,

$$p(x) = \frac{N!}{x!(N-x)!} \left(\frac{\lambda}{N}\right)^x \left(1 - \frac{\lambda}{N}\right)^{N-x}$$

But we have $\frac{N!}{(N-x)!}$ is around N^x for $x \ll N$. We also have $\lim_{N\to\infty}(1-\frac{\lambda}{N})^N = e^{-\lambda}$ so is $\lim_{N\to\infty}(1-\frac{\lambda}{N})^{N-x} = e^{-\lambda}$. This gives us,

$$p(x) = \frac{N^x}{x!}\left(\frac{\lambda}{N}\right)^x e^{-\lambda} = e^{-\lambda}\frac{\lambda^x}{x!}$$

This is called a **Poisson distribution** ◇

Note. Poisson is from $0 \to \infty$

If $p(x)$ is the p.m.f of a random variable, then, we need $p(x) \geq 0, \forall x$ and $\sum_x p(x) = 1$

Example. For binomial we have,

$$p(x) = \binom{N}{x} p^x (1-p)^{N-x}$$

Now,

$$\begin{aligned}\sum_{x=0}^{N} p(x) &= \sum_{x=0}^{N}\binom{N}{x} p^x(1-p)^{N-x} \\ &= (p+(1-p))^N \\ &= 1^N = 1\end{aligned}$$

◇

Example. For Poisson we have,

$$\begin{aligned}\sum_{x=0}^{\infty} p(x) &= \sum_{x=0}^{\infty}\frac{\lambda^x}{x!}e^{-\lambda} \\ &= e^{\lambda}e^{-\lambda} = 1\end{aligned}$$

◇

2.3 Mean of a random variable

Also called average, expectation etc. Here $\mathbb{E}(x)$ is the expectation of X.

Example. With probability p say you get 1 and with $(1-p)$ you get 0. Here the average after N flips is,

$$1 \cdot \frac{\#1}{N} + 0 \cdot \frac{\#0}{N} = 1 \cdot p(1) + 0p(0) = p$$

◇

Definition. If X is a discrete random variable with p.m.f $p(x)$ then,

$$\mathbb{E}(X) = \sum_x x \cdot p(x)$$

Note. Here x is all possible values of X.

Example. For Bernoulli r.v.,

$$\mathbb{E}(X) = p$$

◇

Example. For Binomial N, p we have,

$$\mathbb{E}(X) = \sum_{x=0}^{N} x\binom{N}{x} p^x(1-p)^{N-x}$$

First we have $x\binom{N}{x} = \frac{N!}{(x-1)!(N-x!)} = \frac{(N-1)!(N)}{(x-1)!(N-x)!} = N\binom{N-1}{x-1}$. So,

$$N\sum_{x=1}^{N}\binom{N-1}{x-1}p^x(1-p)^{N-x} = Np\sum_{x=1}^{N}\binom{N-1}{x-1}p^{x-1}(1-p)^{N-x}$$

Taking $y = x-1$ we have,

$$Np\sum_{y=0}^{N-1}\binom{N-1}{y}p^y(1-p)^{N-1-y} = Np$$

$\diamond$

Remark. Intuitively if $Y_1, \ldots, Y_n$ are independent Bernoulli r.v. with parameter p. Then $\sum_{i=1}^{N} Y_i = X$ binomial and $\mathbb{E}(\sum_{i=1}^{N} Y_i) = \sum_{i=1}^{N}\mathbb{E}(Y_i) = Np$

Example. For Geometric r.v. we have,

$$p(x) = p(1-p)^{x-1}$$

So,

$$\sum_{x=0}^{\infty} xp(x) = p\sum_{x=0}^{\infty} x(1-p)^{x-1}$$

We see that $x(1-p)^{x-1} = \frac{-d}{dp}(1-p)^x$ so,

$$\begin{aligned}
\sum_{x=1}^{\infty} x(1-p)^{x-1} &= -\sum_{x=1}\frac{d}{dp}(1-p)^x \\
&= -\frac{d}{dp}\sum_{x=1}^{\infty}(1-p)^x \\
&= -(1-p)\frac{d}{dp}\sum_{x=0}^{\infty}(1-p)^x \\
&= -\frac{d}{dp}(\frac{1}{p}-1) \\
&= -1\times -\frac{1}{p^2} = \frac{1}{p^2}
\end{aligned}$$

So we have,

$$\mathbb{E}(X) = p\sum_{n=1}^{\infty} x(1-p)^{x-1} = p\frac{1}{p^2} = \frac{1}{p}$$

$\diamond$

Example. For Poisson we have,

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

$\diamond$

Example. For Hypergeometric we have,

$$p(x) = \frac{\binom{N}{x}\binom{M}{n-x}}{\binom{M+N}{n}}$$

In this example we have N red balls and M blue balls and X is the r.v. of number of reds by taking x balls. Intuitively we get $\mathbb{E}(X) = np$ where $p = \frac{N}{N+M}$ ◇

Remark. The point here is that extracting the red ball at the third extraction and the first extraction will be the same (at the beginning that is, even though given information about the first 2 extractions the third will have different probability). But in the beginning we don't have any more information and there is no reason to think that the first or second extraction is better than the later ones.

2.4 Review

1. A discrete r.v. $X : \Omega \to \mathbb{R}$ where $X(\Omega)$ is finite or countable and we have $X^{-1}(x) \in \mathscr{F}$.
2. The probability mass function (pmf) is defined as,

$$p_X(x) = \mathbb{P}(X = x)$$

Example. We defined the following random variables,

1. Bernoulli r.v. $X \sim (p)$, $p_X(1) = p$, $p_X(0) = 1 - p$.
2. Binomial r.v. $X \sim (n, p)$, $p_X(k) = \binom{n}{k}p^k(1-p)^{n-k}$, $k = 0, 1, \ldots, n$
3. Hypergeometric r.v. $X \sim (N, K, n)$, $p_X(k) = \frac{\binom{K}{k}\binom{N-K}{n-k}}{\binom{N}{n}}$, $k = 0, 1, \ldots, \min\{n, K\}$.
4. Geometric r.v. $X \sim (p)$, $p_X(k) = (1-p)^{k-1}p$, $k = 1, 2, \ldots$
5. Poisson r.v. $X \sim (\lambda)$, $p_X(k) = e^{-\lambda}\frac{\lambda^k}{k!}$, $k = 0, 1, 2, \ldots$

◇

We also have $\mathbb{E}(X) = \sum_x x p_X(x)$ and got,

Example. For Bernoulli, $\mathbb{E}(x) = p$,
For Binomial, $\mathbb{E}(X) = np$,
For Hypergeometric, $\mathbb{E}(X) = \frac{N}{N+M}n$
For Geometric, $\mathbb{E}(X) = \frac{1}{p}$,
For Poisson, $\mathbb{E}(X) = \lambda$. ◇

Consider a r.v. Y such that $ImY = \{a, b\}$ and,

$$\mathbb{P}(X = b) = p$$

And have,

$$X = \frac{Y - a}{b - a}$$

Then X takes two values 0 and 1 with probabilities $1 - p$ and p. So,

$$\mathbb{P}(X = 1) = \mathbb{P}(Y = b) = p$$

So X is a bernoulli r.v. with parameter p and,

$$X = \frac{Y - a}{b - a} \Rightarrow Y = a + (b - a)X$$

So any function that takes two values can be written as a linear functions of a Bernoulli r.v. And,

$$
\begin{aligned}
\mathbb{E}(Y) &= (1-p)a + pb \\
&= a + (b-a)p
\end{aligned}
$$

We can also say that,

$$
\begin{aligned}
\mathbb{E}(Y) &= \mathbb{E}(a + (b-a)X) \\
&= a + (b-a)\mathbb{E}(X) \\
&= a + (b-a)p
\end{aligned}
$$

Theorem 2.1. So if X is a r.v. and $a, b \in \mathbb{R}$ then,

$$\mathbb{E}(a + bX) = a + bE(X)$$

Proof. We have $Y = a+bX$. The possible values of Y are $\{y : y = a+bx, x \text{ is a possible value for } x\}$
Here,

$$p_y(y) = p_x(x) \text{ where } y = a + bx$$

So,

$$
\begin{aligned}
\mathbb{E}(Y) &= \sum_y y p_Y(y) \\
&= \sum_x (a + bx) p_X(x) \\
&= a \sum_x p_X(x) + b \sum_x x p_X(x) \\
&= a + bE(X)
\end{aligned}
$$

□

Example. If X is a r.v. then $Y = X^2$. We have,

$$\mathbb{P}(Y = y) = \mathbb{P}(X = \pm\sqrt{y})$$

so,

$$p_Y(y) = p_X(\sqrt{y}) + p_X(-\sqrt{y})$$

Then we have,

$$
\begin{aligned}
E(Y) &= \sum_y y p_Y(y) \\
&= \sum_y y(p_X(\sqrt{y}) + p_X(-\sqrt{y})) \\
&= \sum_x x^2 p_X(x) \\
&= E(X^2)
\end{aligned}
$$

Thus we have $\mathbb{E}(X^2) = \sum_x x^2 p_X(x)$ ◇

Theorem 2.2. Let X be a r.v. and $f : \mathbb{R} \to \mathbb{R}$. Then,

$$Y = f(x) \text{ is a r.v.}$$

$$\mathbb{P}(Y = y) = \sum_{x:f(x)=y} \mathbb{P}(X = x)$$

Moreover,

$$\mathbb{E}(f(x)) = \mathbb{E}(Y) = \sum_x f(x) p_X(x)$$

Remark. If $A_x = \{\omega | X(\omega) = x\}$ we have $A_x \in \mathscr{F}$ then,

$$\{\omega | Y(\omega) = y\} = \bigcup_{x:f(x)=y} A_x$$

Example. Take $\mathbb{E}(X^n) = m_n(X)$ is the $n'th$ moment of X.

We would also like to know what the width of the distribution is. A naive approach is $\mathbb{E}(X - \mathbb{E}(X)) = 0$ which is useless. But we can define,

$$V(X) = \mathbb{E}((X - \mathbb{E}(X))^2)$$

which is the average squared distance from the average.
And we call,

$$\sigma_X^2 = V(X) \text{ where } \sigma \text{ is the standard deviation}$$

◇

Example. Consider for bernoulli,

$$\mathbb{E}(X^2) = 1p + 0(1 - p) = p$$

But,

$$\begin{aligned}\mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}((X - p)^2) &= (1 - p)^2 p + (0 - p)^2 (1 - p) \\ &= p(1 - p)\end{aligned}$$

◇

Remark. Take $\mathbb{E}(X) = m$ then we can simplify $V(X)$ as follows,

$$\begin{aligned}V(X) &= \mathbb{E}((X - m)^2) \\ &= \mathbb{E}(X^2 - 2mX + m^2) \\ &= \mathbb{E}(X^2) - 2m\mathbb{E}(X) + m^2 \\ &= \mathbb{E}(X^2) - 2m^2 + m^2 \\ &= \mathbb{E}(X^2) - m^2 \\ &= \mathbb{E}(X^2) - (\mathbb{E}(X))^2\end{aligned}$$

So using this formula for bernoulli, we have,

$$V(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 = p - p^2 = p(1 - p)$$

Theorem 2.3. If X is a random variable then $\mathbb{E}(X^2) \geq \mathbb{E}(X)^2$.

Proof. We have $V(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 \geq 0$. Thus, $\mathbb{E}(X^2) \geq (\mathbb{E}(X))^2$. $\square$

2.5 Variance

Definition. If X is a discrete r.v. then, variance of X is defined as

$$Var(X) = E[(X - E[X])^2]$$

Remark. But we can also write it as,

$$\begin{aligned} Var(X) &= E[(X - E[X])^2] \\ &= E[X^2] - E[X]^2 \end{aligned}$$

Example. If we have a **bernoulli r.v**. we have,

$$\begin{aligned} E(X) &= p \\ Var(X) &= p(1-p) \end{aligned}$$

We have $V(X)$ is maximal if $p = \frac{1}{2}$ and zero if $p = 0$ or $p = 1$ $\diamond$

Example. For **Binomial r.v.**, we have,

$$\begin{aligned} E(X) &= np \\ Var(X) &= np(1-p) \end{aligned}$$

We have,

$$\begin{aligned} E(X(X-1)) &= \sum_{n=0}^{\infty} x(x-1)\binom{n}{x} p^x (1-p)^{n-x} \\ &= n(n-1)p^2 \end{aligned}$$

Now we have $Var(X) = E(X(X-1)) + E(X) - E(X)^2$ which is,

$$= np(1-p)$$

$\diamond$

Example. Consider **Hypergeometric r.v.**, with parameters N, M, n. We have,

$$p(x) = \frac{\binom{N}{x}\binom{M}{n-x}}{\binom{M+N}{n}}$$

We know that,

$$E(X) = np$$

and that

$$V(X) = np(1-p)\frac{N+M-n}{N+M-1}$$

$\diamond$

Example. For **Poisson r.v.**, we have,

$$p(x) = \lambda^x \frac{1}{x!} e^{-\lambda}$$

and we can do,

$$\begin{aligned} E(X(X-1)) &= \sum_{x=0}^{\infty} x(x-1)\lambda^x \frac{1}{x!} e^{-\lambda} \\ &= \lambda^2 \end{aligned}$$

Now we have $Var(X) = E(X(X-1)) + E(X) - E(X)^2$ which is,

$$\begin{aligned} Var(X) &= \lambda^2 + \lambda - \lambda^2 \\ &= \lambda \end{aligned}$$

◇

Remark. Poisson variance is also similar to binomial variance, where $n \to \infty$.
Consider,
Example.

$$p(x) = \frac{C}{x^2} \qquad \text{for } x \geq \text{N}$$

We know that $\sum_{x=1}^{\infty} \frac{1}{x^2}$ converges to some k. So we have,

$$C = \frac{1}{k}$$

But we have the $E(X)$ doesn't exist as,

$$\begin{aligned} E(X) &= \sum_{x=N}^{\infty} x \frac{C}{x^2} \\ &= C \sum_{x=N}^{\infty} \frac{1}{x} \end{aligned}$$

But the sum is not divergent so we can say that $E(X) = +\infty$

Here for X the expected value is dominated by large values of X which are rare but have a large contribution to the expected value. ◇

2.6 Conditional Expectation

Definition. We know that if A is an event with $\mathbb{P}(A) > 0$ then $\mathbb{Q}(B) = \mathbb{P}(B \mid A)$ is a probability. So if X is a random variable then,

$$\begin{aligned} E_Q(X) &= \sum_x \mathbb{P}(X = x \mid A) \\ &= E(X \mid A) \end{aligned}$$

which is called te conditional expectation of X given A.

Theorem 2.4. Let A_i be a partition of Ω such that $\mathbb{P}(A_i) > 0$ for all i. Then,

$$E(X) = \sum_i \mathbb{E}(X \mid A_i)\mathbb{P}(A_i)$$

Proof.

$$\begin{aligned} E(X) &= \sum_i x\mathbb{P}(X = x) \\ &= \sum_i x \sum_j \mathbb{P}(X = x \mid A_j)\mathbb{P}(A_j) \\ &= \sum_j \mathbb{P}(A_j) \sum_x x\mathbb{P}(X = x \mid A_j) \\ &= \sum_j \mathbb{E}(X \mid A_j)\mathbb{P}(A_j) \end{aligned}$$

□

Example. Consider if 10 percent of the population is sick and 90 percent is not. Now take n individual and X is the r.v denoting the number of sick people. As N is large, we can approximate X by a binomial r.v. with parameters $n, p = 0.1$. Now we do a test and consider,

$$\mathbb{P}(P \mid Notsick) = 0.1, \mathbb{P}(NP \mid Sick) = 0$$

Take the n extracted people and test them. Let Y be the number of positive tests. Here Y is also binomial. Now we have $\mathbb{P}(P) = \mathbb{P}(P \mid S)\mathbb{P}(S) + \mathbb{P}(P \mid NS)\mathbb{P}(NS) = 0.1 \times 0.9 + 1 \times 0.1 = 0.19$. Now we have,

$$E(Y) = n \times 0.19$$

◇

Example. Flip a coin and look at the initial string of heads. Let X be the length of this string. We have arrival of first head as $\frac{1}{p} - 1$. Now we have, for tails at first flip $\frac{1}{q} - 1$ so we have,

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

Another way to do this is,

$$\begin{aligned} E(X) &= E(X \mid H)\mathbb{P}(H) + E(X \mid T)\mathbb{P}(T) \\ &= p^{x-1}qp + q^{x-1}pq \end{aligned}$$

which gives us the same answer.

◇

Remark. The logic in the first method is that arrival of first head means that the rest before it is tails i.e. a continuous run of tails. So as $1/p$ includes the heads we do $\frac{1}{p} - 1$ which represent the average length of a run of tails. We get $1/p$ as that is the mean of the geometric random variable.

Another way of doing this is to define X as the number of continuous run of heads. So $\mathbb{P}(X = x) = p^x q$ and finding the expected value of this gives us the desired answer.

Chapter 3

Multivariate discrete distribution and Independence

Example. Roll 2 dice (blue and red). So,

$$\Omega = \{(x, y) \mid 1 \leq x, y \leq 6\}$$

We have X outcome of the first dice and Y the outcome of the second, with

$$X(\Omega) = Y(\Omega) = \{1, \ldots, 6\}$$

and,

$$p(x, y) = \mathbb{P}(X = x, Y = y) = \frac{1}{36}$$

So here $p(x, y)$ is called the *joint probability mass function* of the r.v. X and Y.

We can say $\mathbb{P}(X = x) = \sum_y \mathbb{P}(X = x \mid Y = y) = p_X(x)$ which is called the marginal over X of $p(x, y)$
Similarly, $\mathbb{P}(Y = y) = \sum_x \mathbb{P}(Y = y \mid X = x) = p_Y(y$ which is the marginal over Y of $p(x, y)$

In this case of the dice we have,

$$p_X(x) = p_Y(y) = \frac{1}{6} \quad \forall x, y$$

$\diamond$

Example. Consider the above example but now let $Z_+ = X + Y$ and $Z_- = X - Y$. So we have,

$$Z_+ = \{2, \ldots, 12\} \quad \text{and} \quad Z_- = \{-5, \ldots, 5\}$$

Now we can say that,

$$\begin{aligned}
Z_+ = 2 &\Rightarrow Z_- = 0 \\
Z_+ = 3 &\Rightarrow Z_- = \{1, -2\} \\
Z_+ = 4 &\Rightarrow Z_- = \{-2, 0, 2\} \\
&\vdots
\end{aligned}$$

We can ask $\mathbb{P}(Z_+ = 4 \text{ and } Z_- = -2) = \frac{1}{36}$ as there is only one possibility which is when $X = 1, Y = 3$.

Notice that the set of possible values of Z_- is like a rhomboid i.e. it increases linearly until 7 and then decreases linearly.

We can ask $\mathbb{P}(Z_+ = 4) = \sum_z \mathbb{P}(Z_+ = 4 \text{ and } Z_i = z) = \frac{3}{36} = \frac{1}{12}$

$\diamond$

Definition. If x and y are r.v. over Ω then,

$$p(x, y) = \mathbb{P}(X = x \text{ and } Y = y)$$

is called the joint p.m.f of X and Y. And we must have,

$$\begin{aligned} p(x, y) &\geq 0 \\ \sum_{x,y} p(x, y) &= 1 \end{aligned}$$

3.1 Expected Values

Given X, Y we want $\mathbb{E}(X)$. We have,

$$\begin{aligned} \mathbb{E}(X) &= \sum_x x p_X(x) \\ &= \sum_x x \sum_y \mathbb{P}(X = x \text{ and } Y = y) \\ &= \sum_{x,y} x \mathbb{P}(X = x \text{ and } Y = y) \end{aligned}$$

Similarly,

$$\mathbb{E}(Y) = \sum_{x,y} y \mathbb{P}(X = x \text{ and } Y = y)$$

Given (X, Y), I can think of X as a function of (X, Y). Take a less simple function,

$$Z = aX + bY \quad a, b \in \mathbb{R}$$

We can ask $\mathbb{E}(Z)$ we have,

$$\begin{aligned} \mathbb{E}(Z) &= \sum_z \mathbb{P}(Z = z) \\ &= \sum_{x,y} (ax + by) \mathbb{P}(X = x \text{ and } Y = y) \\ &= a \sum_{x,y} x \mathbb{P}(X = x \text{ and } Y = y) + b \sum_{x,y} y \mathbb{P}(X = x \text{ and } Y = y) \\ &= a\mathbb{E}(X) + b\mathbb{E}(Y) \end{aligned}$$

Corollary 3.1.

$$\mathbb{E}(aX + bY + c) = a\mathbb{E}(X) + b\mathbb{E}(Y) + c$$

Consider r.v X, Y and $g : \mathbb{R}^2 \to \mathbb{R}$ so that,

$$Z = g(X, Y)$$

Since X, Y are r.v we can say,

$$\{\omega \mid X(\omega) = x \text{ and } Y(\omega) = y\} = \{\omega \mid X(\omega) = x\} \cap \{\omega \mid Y(\omega) = y\} \in \mathscr{F}$$

We have,

$$\begin{aligned} \mathbb{E}(Z) &= \mathbb{E}(g(X, Y)) \\ &= \sum_{x,y} g(x, y)\mathbb{P}(X = x \text{ and } Y = y) \end{aligned}$$

3.2 Covariance

Definition. Covariance is defined as,

$$\begin{aligned} Cov(X, Y) &= \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y))) \\ &= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) \end{aligned}$$

Note. The covariance of X with itself is just the variance of X.
Note. We can used this to measure how two r.v are correlated.
We can define the correlatoin coefficient as,

$$\rho = \frac{Cov(X, Y)}{\sigma_X \sigma_Y}$$

Where $\sigma_X^2 = Var(X)$ and $\sigma_Y^2 = Var(Y)$ and,

$$-1 \leq \rho_{x,y} \leq 1$$

We also have,

$$\begin{aligned} \rho_{X,Y} &= 1 \quad X = aY \quad a > 0 \\ \rho_{X,Y} &= -1 \quad X = bY \quad b < 0 \\ \rho_{X,Y} &= 0 \quad \text{we say they are uncorrelated} \end{aligned}$$

3.3 Independence

Definition. If X, Y are independent then if I know the value of X it gives me no information on the value of Y. So we have,

$$\mathbb{P}(Y = y \mid X = x) = \mathbb{P}(Y = y)$$

or that,

$$\mathbb{P}(Y = y \text{ and } X = x) = \mathbb{P}(X = x)\mathbb{P}(Y = y) \quad \forall x, y$$

Theorem 3.2. If X, Y are independent then we have $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$

Proof. We have,

$$\begin{aligned}
\mathbb{E}(XY) &= \sum_{x,y} xy\mathbb{P}(X = x \text{ and } Y = y) \\
&= \sum_{x,y} xy\mathbb{P}(X = x)\mathbb{P}(y = y) \\
&= \sum_{x,y} x\mathbb{P}(X = x) \sum_{y} y\mathbb{P}(y = y) \\
&= \mathbb{E}(X)\mathbb{E}(Y)
\end{aligned}$$

$\square$

Remark. Which also means that we have $(XY) = 0$. However,

$$Cov(X, Y) = 0 \not\Rightarrow \text{ independence}$$

For instance consider we have possible values only $\{(0, 1), (1, 0), (-1, 0), (0, -1)\}$. So X can take value -1 with probability $\frac{1}{4}$, 0 with $\frac{1}{2}$ and 1 with $\frac{1}{4}$. We have $\mathbb{E}(X) = 0, \mathbb{E}(Y) = 0$ and $XY = 0$. So we get $Cov(X, Y) = 0$. But if we know that value of X then we know the possible value of Y. More specifically we have,

$$\mathbb{P}(X = 0) = \mathbb{P}(Y = 0) = \frac{1}{2} \quad \text{but} \quad \mathbb{P}(X = 0 \text{ and } Y = 0) = 0$$

Here X, Y are not independent but $Cov(X, Y) = 0$.

Remark. So if X, Y are independent and g, f are functions then we get,

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

Definition. In general given $X_1, \ldots, X_n$ then the joint p.m.f is,

$$\mathbb{P}(X_1 = x_1 \ldots X_n = x_n)$$

Remark. Independence means that $\mathbb{P}(X_1 = x_1 \;\&\; \ldots \;\&\; X_n = x_n) = \prod_{i=1}^{n} \mathbb{P}(X_i = x_i)$

We can further show this is true for any subset of X by summing over X_i to exclude X_i.

We can recognize independence if, $p(x, y) = f(x)g(y)$. For instnace,

$$\begin{aligned}
p(x, y) &= \frac{e^{-\lambda-\mu}}{x!y!}\mu^x\lambda^y \\
&= \left(\frac{e^{-\lambda-\mu}}{x!}\mu^x\right)\left(\frac{\lambda^y}{y!}\right)
\end{aligned}$$

So we're able to write $p(x, y)$ as the product of two functions dependent on x, y.

We know that $V(X) = \mathbb{E}((X - E(X)^2)$. So we look at $V(X + Y)$ and get,

Theorem 3.3. If X and Y are independent then we have,

$$V(X + Y) = V(X) + V(Y)$$

Proof.

$$\begin{aligned}
V(X+Y) &= \mathbb{E}(((X+Y)-E(X+y))^2) \\
&= \mathbb{E}((X+Y-\mu_x-\mu_y)^2)) \\
&= \mathbb{E}((X-\mu_x)^2+(Y-\mu_y)^2+2(X-\mu_x)(Y-\mu_x)) \\
&= \mathbb{E}((X-\mu_x)^2)+\mathbb{E}((Y-\mu_y)^2)+\mathbb{E}(2(X-\mu_x)(Y-\mu_x))) \\
&= Var(X)+(Y)+2Cov(X,Y)
\end{aligned}$$

So we have $V(X+Y)=V(X)+V(Y)$ if we have $Cov(X,Y)=0$. But if X,Y are independent then we have $Cov(X,Y)=0$ □

Remark. Note that $V(X+Y)=V(X)+V(Y)$ does **NOT** mean that X,Y are independent.

Remark. Let $X_1,\dots,X_n$ be independent Bernoulli r.v. of parameter p. Consider n independent coin flips so,

$$Y=\sum_{i=1}^{n} X_i$$

Y is binomial n,p. So we have,

$$\mathbb{E}(Y)=\sum_i \mathbb{E}(X_i)=np$$

Now for variance of Y we have,

$$V(Y)=\sum_i V(X_i)=np(1-p)$$

as each X_i is independent from each other and variance of X_i is $(1-p)$

Given $V(aX)$ we have $V(aX)=\mathbb{E}(a^2X^2)-\mathbb{E}(aX)^2=a^2V(X)$. This also gives us standard deviation $\sigma_{aX}=|a|\sigma_X$.

We also have $V(X+a)=V(X)$ as,

$$\begin{aligned}
\mathbb{E}(X+a) &= \mathbb{E}(X)+a \\
(X+a)-\mathbb{E}(X+a) &= X-\mathbb{E}(X)
\end{aligned}$$

Remark. If variance gives an idea of spread of the distribution then adding a constant value to a r.v. X should not change the spread.

Remark. We can also see it by looking at a a constant as independent from X and then use linearity of variance when independence is true to separate it out.

3.4 Sum of random variables

Take X,Y with $p(x,y)$. Now consider,

$$Z=X+Y$$

We have,

$$\begin{aligned}
\mathbb{P}(Z=z) &= \sum_x \mathbb{P}(X=x \;\&\; Y=z-x) \\
&= \sum_x p(x,z-x)
\end{aligned}$$

If X and Y are independent then,

$$\mathbb{P}(Z=z)=\sum_x p_X(x)p_Y(z-x)$$

This is called the convolution.

Example. If X and Y are Poisson, then with μ, ν we have,

$$p_X(x) = \frac{\mu^x}{x!}e^{-\mu} \quad p_Y(y) = \frac{\nu^y}{y!}e^{-nu}$$

Now consider $Z = X + Y$ then,

$$\begin{aligned}
\mathbb{P}(Z = z) &= \sum_{x=0}^{z} p_X(x)p_Y(z-x) \\
&= \sum_{x=0}^{z} \frac{\mu^x \nu^{z-x}}{x!(z-x)!}e^{-(\mu+\nu)} \\
&= \sum_{x=0}^{z} \frac{z!}{z!}\frac{\mu^x \nu^{z-x}}{x!(z-x)!}e^{-(\mu+\nu)} \\
&= \frac{1}{z!}\sum_{x=0}^{z} \binom{z}{x}\mu^x \nu^{z-x} e^{-(\mu+\nu)} \\
&= \frac{1}{z!}(\mu+\nu)^z e^{-(\mu+\nu)}
\end{aligned}$$

$\diamond$

Exercise. If X is binomial n, p and Y is binomial m, p then $X + Y$ is binomial $n + m, p$

Remark. If X is binomial n, p_1 and Y is binomial n, p_2 where $p_1 \neq p_2$. Then $X+Y$ is not binomial. But we have,

$$\mathbb{E}(X + Y) = np_1 + mp_2$$

and,

$$V(X + Y) = np_1(1 - p_1) + mp_2(1 - p_2)$$

3.5 Indicator Function

Definition. Indicator function is a r.v $1_A(\omega)$ such that,

$$1_A(\omega) = \begin{cases} 1 & \omega \in A \\ 0 & \omega \notin A \end{cases}$$

We have the follwoing,

$$\begin{aligned}
1_A^c(\omega) &= 1 - 1_A(\omega) \\
1_{A\cup B}(\omega) &= 1_A(\omega)1_B(\omega) \\
1_{A\cap B}(\omega) &= 1_{(A^c\cap B^c)^c}(\omega) = 1 - (1 - 1_A(\omega))(1 - 1_B(\omega)) \\
&= 1_A(\omega) + 1_B(\omega) - 1_A(\omega)1_B(\omega)
\end{aligned}$$

Now we have,

$$\mathbb{E}(1_{A\cup B}) = \mathbb{E}(1_A) + \mathbb{E}(1_B) - \mathbb{E}(1_A 1_B)$$

this tells us that,

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

In general,

$$1_{\bigcup_i A_i}(\omega) = 1 - \prod_i (1 - 1_{A i}(\omega))$$

We can write this as,

$$\sum_{I\in\{1,\ldots,n\}} (-1)^{|I|+1}\mathbb{P}\left(\bigcap_{i\in I} A_i\right)$$

Example. Consider a gas station on a highway and N_1 a poisson r.v λ_1 representing a car that stops and needs service on top of gas. Let N_2 be poisson λ_2, for a car that stops and do not need service.
So we have,

$$N_1 + N_2 = \text{ Poisson with } \lambda_1 + \lambda_2$$

We can describe it as follows as well,

N is a Poisson with prob λ for every car that arrives with probability p of requiring service and $1-p$ of not requiring service.

Now let N_1 is the number of car that requires service. We have $N_1$is a Poisson r.v with parameter $p\lambda$. We can compute this by taking $\mathbb{P}(N_1 = n) = \sum_{m \geq n} \mathbb{P}(N_1 = n \mid N = m)\mathbb{P}(N \mid m)$
The first term is binomial with m, p and second term is a poisson. So this will give us,

$$\sum_{n \geq m} \binom{m}{n} p^n (1-p)^{m-n} \frac{\lambda^m}{m!} e^{-\lambda} = \frac{(\lambda p)^n}{n!} e^{-\lambda p}$$

But an easier way to do this is by assuming that N is actually a binomial with large n and $p = \frac{\lambda}{n}$. But if we have a binomial then at each time step a car arrives with probability $\frac{\lambda}{n}$ and probability is p that car needs service so we have $\frac{p\lambda}{n}$ as probability of car arriving and needing service at each time step.

So consider a binomial with n flips and $\frac{p\lambda}{n}$ which is a Poisson with parameter $p\lambda$

So if N_2 is the number of cars that do not need service it is a Poisson with $(1-p)\lambda$ $\diamond$

Remark. Point here is that sum of two independent Poisson is Poisson.