4) Discrete Random Variables:

Defn A *random variable* is a quantity of interest that depends on outcome of a probability experiment.

Example: 4.1 Consider the following bet: you roll a fair die and win

(a) win £2 if the outcome is 5 or 6

(b) lose £1 if the outcome is 1, 2 or 3

(c) win or lose nothing if outcome is 4.

Sample space is $\Omega = \{1, 2, 3, 4, 5, 6\}$

Denoting losses as negative gains, we can represent gains from this bet as a function

$X: \Omega \to \mathbb{R}$ defined as

$$X(\omega) = \begin{cases} -1 & \text{if } \omega = 1, 2, 3 \\ 0 & \text{if } \omega = 4 \\ 2 & 2 \ \ \omega = 5, 6 \end{cases}$$

The amount gain is a random variable

Defn 4.2: A **random variable** is a function from a sample space into the real numbers $\mathbb{R}$.

$$X : \Omega \to \mathbb{R}$$

Discrete random variables are random variables that take only countably many values.

The random variable in 4.1 is a discrete random variable with image

$$X(\Omega) = \{-1, 0, 2\}$$

$X : \Omega \to \mathbb{R}$ is essentially a function that maps values from Ω to real numbers $\mathbb{R}$.

$X(\Omega)$ is usually the image of X.

Defn: The image is the set of values in codomain $\mathbb{R}$ (i.e. a subset of $\mathbb{R}$) that the function takes as we plug in all elements of Ω

$$X(\Omega) = \{X(\omega) \mid \omega \in \Omega\}$$

The different values that a random variable can take define different events.

In Example 4.1,

$X^{-1}(-1) = \{w \in \Omega \mid X(w) = -1\} = \{X = -1\} = \{1, 2, 3\}$

$X^{-1}(0) = \{w \in \Omega \mid X(w) = 0\} = \{X = 0\} = \{4\}$

$X^{-1}(2) = \{w \in \Omega \mid X(w) = 2\} = \{X = 2\} = \{5, 6\}$

These are preimages.

Defn: $X^{-1}(1)$ is basically a preimage basically stating that $-1 \in \mathbb{R}$, we can ask which elements of Ω when we plug in to X and end up in $\{-1\}$ which is a subset of $\mathbb{R}$. So

$$X^{-1}(-1) = \{w \in \Omega \mid X(w) = -1\}$$

Let $X \subseteq \mathbb{R}$, X is a subset of $\mathbb{R}$.

$$\boxed{X^{-1}(x) = \{w \in \Omega \mid X(w) \in X\}}$$

These events form a partition of Ω.
We will often be interested in the probability of these events.

Notation: We just write $P(X=2)$ to denote $P(\{X=2\})$ or $P(\{w\in\Omega \mid X(w)=2\})$.

4.2 The probability distribution of a discrete random variable.

Defn 4.4: The probability mass function of discrete random variable X is the function:

$$\boxed{p_X : \mathbb{R} \to \mathbb{R} \text{ defined by } p_X(x) = P(X=x)}$$

Notation: Capital letters for random variables
Lower case letters for real numbers.

In example 4.1 the probability mass fn is given by

$$p_X(x) = \begin{cases} 1/2 & \text{if } x = -1 \\ 1/6 & \text{if } x = 0 \\ 1/3 & \text{if } x = 2 \\ 0 & \text{if } x \notin \{-1, 0, 2\} \end{cases}$$

↳ always write down 0 possibility

The explanation is given below:

So the sample space is $\Omega = \{1, 2, 3, 4, 5, 6\}$ and the random variable X is a function such that

$$X : \Omega \to \mathbb{R},$$

$$X(\omega) = \begin{cases} -1 & \text{if } \omega = 1, 2, 3 \\ 0 & \text{if } \omega = 4 \\ 2 & \text{if } \omega = 5, 6 \end{cases}$$

and that $\omega \in \Omega$

As we can see, for $X(\omega) = -1$, there are three possible outcomes out of 6, i.e.

$$X^{-1}(-1) = \{1, 2, 3\}.$$

Hence

$$\boxed{p_X(x) = P(X=x) = P(\{X=x\}) = P(X^{-1}(x))}$$

Then for $X = -1$,

$$p_X(-1) = P(X=-1) = P(\{X=-1\}) = P(X^{-1}(1))$$

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

$$= 1/2$$

By probability laws and theorems found in Chapter 2.

Similar explain for $X = 0$ and $X = 2$.

$X : \Omega \to \mathbb{R}$ is a function that maps from Ω to $\mathbb{R}$.
But probabilities are only defined on the image of $X(\Omega)$ where $X(\Omega) \subseteq \mathbb{R}$.
For all other values of $\mathbb{R}$, probabilities are 0.
Hence probability is 0 for all $x \notin X(\Omega) \subseteq \mathbb{R}$.

For example 4.1,

$$X(\Omega) = \{-1, 0, 2\}$$

Hence

$$p_X(x) = 0 \quad \text{for} \quad x \notin X(\Omega) = \{-1, 0, 2\}$$

Theorem 4.5: Let X be a random variable with countable image $I = X(\Omega)$.
Then its probability mass function p_X satisfies

(m1) $\quad p_X(x) \geq 0 \;\; \forall x \in \mathbb{R}$ and $p_X(x) = 0 \;\; \forall x \notin I$

(m2) $$\sum_{x \in I} p_X(x) = 1$$

proof: For (m_1):

$p_X(x) = P(X = x)$ and hence by $(P1)$,

$P(X = x) \in [0,1]$ and therefore

$$P(X = x) = p_X(x) \geq 0$$

For $(m2)$:

$$\sum_{x \in I} p_X(x) = \boxed{\sum_{x \in I} P(X = x) = P\Big(\bigcup_{x \in I} \{X = x\}\Big)}$$

$$= P(\Omega) = 1$$

☆,
explanation next page

$\blacksquare$

$\star_1$ Applying (P3). We have mentioned that all the X^{-1}'s (preimages) partion Ω hence form disjoint events; $x, y \in I$; $X^{-1}(y) \cap X^{-1}(x) = \phi$

Hence $\sum_{x \in I} P(X=x)$ can be the union of all $X=x = \{X=x\}$ on which probability function is applied where
$$\{X=x\} = \{w \in \Omega \mid X(w) = x\}$$

All $\{X=x\} = \{w \in \Omega \mid X(w)=x\} = X^{-1}$ are disjoint hence (P3) applied.

Theorem: 4-6 Consider any countable set $I \subset \mathbb{R}$ and any function $p_X : \mathbb{R} \to \mathbb{R}$ satisfying (m1), (m2) above. Then there exists a probability space $(\Omega, \mathcal{F}, P)$ and a discrete random variable $X : \Omega \to \mathbb{R}$ such that p_X is the probability mass function of X.

proof: Simply construct an example of such a probability space and a random variable

Choose $\Omega = I$, $\mathcal{F}$ = set of all subsets of Ω
$= \mathcal{P}(\Omega)$

Also choose P the probability function defined according to Theorem 2.18 by

$$P(\{x\}) = p_X(x) \quad \text{for } x \in \Omega$$

and X the random variable $X : \Omega \to \mathbb{R}$, $X(x) = x$.

Then we have

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

for all $x \in I$ and for $x \notin I$, we have

$$P(X = x) = P(\emptyset) = 0$$

Defn 4.4 implies that p_X is the mass function of X.

$\blacksquare$

Events involving values of a random variable can be assigned probabilities via the distribution function.

Defn 4.7: Let X be the random variable. The distribution function of X is the function

$$\boxed{F_X : \mathbb{R} \to \mathbb{R}}$$

defined by

$$\boxed{F_X(x) = P(X \leq x)} = P\left(X^{-1}\left((-\infty, x]\right)\right)$$

In example 4.1, the distribution function is given by

$$F_X(x) = \begin{cases} 0 & \text{if } x \leq -1 \\ 1/2 & \text{if } -1 \leq x < 0 \\ 1/3 & \text{if } 0 \leq x < 2 \\ 1 & \text{if } 2 \leq x \end{cases}$$

The ranges must cover $\mathbb{R}$

explanation on next page

Basically from defn 4.7: $F_X(x) = P(X \le x)$

- For first case $\underline{x < -1}$: any value you take in $x < -1$, say -2, is not in the image of Ω. $F_X(-2) = P(X \le -2) = 0$ as there is no value in image, i.e. no $x \in X(\Omega)$ s.t $x < -1$. Hence, $F_X(x) = 0$ for $x < -1$. $\{X \le x\}$ for $x < -1$ is $\emptyset$, as no $\{X = x\}$ such that $x < 1$
- For second case: $\underline{-1 \le x < 0}$

$-1 \le x < 0$ preimage $\{X \le x\}$ contains

$$\{X = -1\}$$

So
$$F_X(x) = P(X \le x) = P\{X = -1\}$$
$$= 1/2$$

- For third case: $\underline{0 \le x < 2}$

$0 \le x < 2$ preimage $\{X \le x\}$ contains

$$\{X = -1\} \cup \{X = 0\}$$

since $X \le x$ $\forall$ $0 \le x < 2$ and $\{X = -1\}$ is valid: $-1 \le x$ $\forall$ $0 \le x < 2$

So $F_X(x) = P(X \le x) = P\{X = -1\} + P\{X = 0\}$
$= 1/2 + 1/6 = 2/3$

(by (P3))

- For last case $\underline{2 \le x}$:

$2 \le x$ preimage $\{X \le x\}$ contains $(X \le 2 \le x)$

$$\{X=-1\} \cup \{X=0\} \cup \{X=2\}$$

So $F_X(x) = P(X \le x) = P\{X=-1\} + P\{X=0\} + P\{X=2\}$

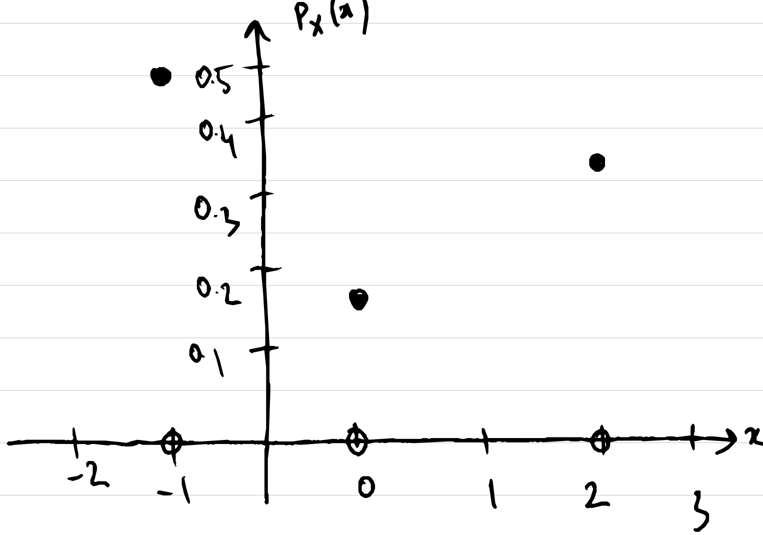
$$= 1/2 + 1/6 + 1/3 = 1.$$

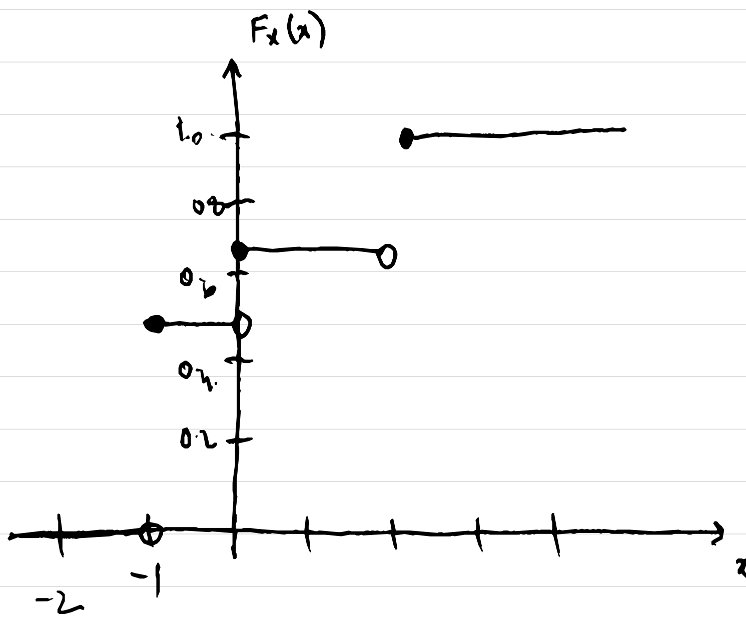
Taking any value $2 \le x$, say 3, $F_X(3) = P(X \le 3)$ and the image $X(\Omega)$ will always contain values for $X \le 3$ for also for any $2 \le x$ for that matter, hence probability is 1.

In general for any discrete random variable $X\{x_0, x_1 \dots\}$ with $x_0 < x_1 < \dots$,

$$\boxed{F_X(x) = P(X \le x) = P\left(\bigcup_{\substack{y \in X(\Omega) \\ y \le x}} \{X=y\}\right) = \sum_{\substack{y \in X(\Omega) \\ y \le x}} P_X(y)}$$

Graphs for $p_x(x)$ and $F_x(x)$ for Ex 4.1





That is distribution function is obtained simply by summing the mass function for all possible values for x.

A different way of saying this is that at every possible value for X, the distribution function F_X jumps by probability of that value

So if x is one of the possible values, then the distribution function at x is larger than the distribution function at left of x by $P_X(x)$

$$\underbrace{P_X(x) = F_X(x) - \lim_{\varepsilon \to 0^+} F_X(x-\varepsilon)}$$

in graph, inclusion only on left side. so to approach it, we must go from right → right limit $\varepsilon \to 0^+$

Note that this allows us to reconstruct the probability mass function $P_X(x)$ from the distribution function F_X.

The theorem below gives the basic properties of distribution functions of random variables

Theorem 4.8: The distribution function F_X of any random variable (continuous and discrete) X satisfies:

(i) $F_X(x)$ is increasing in x

(i.e. if $x \leq y$ then $F_X(x) \leq F_X(y)$)

For $x \leq y$, one that that $F_X(a) \leq F_X(b)$.
An immediate consequence by (P7) since

$\{X \leq x\}$ is contained in $\{X \leq y\}$ $\forall x \leq y$
i.e.
$\{X \leq x\} \subseteq \{X \leq y\}$
So by (P7)

$F_X(x) \leq F_X(y)$ [F is a probability fn so satisfies axioms]

(ii) $\lim_{x \to -\infty} F_X(x) = 0, \quad \lim_{x \to \infty} F_X(x) = 1$

Look at graph to convince.

$$\lim_{x \to \infty} F_X(x) = \lim_{x \to \infty} P(X \leq x) = 1$$

↳ since in limit every event is guaruanteed to happen

$$\lim_{x \to -\infty} F_X(x) = \lim_{x \to -\infty} P(X \leq x) = 0$$

↳ since in limit, no event can have occured.

(iii) $F_X(x)$ is right continuous :

$$\lim_{\varepsilon \to 0^+} F_X(x+\varepsilon) = F_X(x) \text{ for all } x \in \mathbb{R}$$

since all the equalities are on left so we for small bits of graph we go from right to left.

↓

right limit.

4.3 Frequently Used Distributions:

Defn 4.9 **Bernoulli distribution:**

We say that a discrete random variable X has the Bernoulli distribution with parameter p and write

$$\boxed{X \sim Ber(p)}$$

if it only takes values of 0 and 1, i.e.

$$\boxed{X(\Omega) = \{0, 1\}}$$

with

$$\boxed{P(X=1) = p \qquad P(X=0) = 1-p}$$

The mass function of X is

$$\boxed{p_X(x) = \begin{cases} 1-p & \text{if } x=0 \\ p & \text{if } x=1 \\ 0 & \text{if } x \neq 0,1 \end{cases}}$$

The distribution function of X is

$$F_X(x) = P(X \le x) = \begin{cases} 0 & \text{if } x < 0 \\ 1-p & \text{if } 0 \le x < 1 \\ 1 & \text{if } 1 \le x \end{cases}$$

Bernoulli distribution only has 2 possibilities often referred to as "success or failure."
↓
encoded by 0 or 1.

Example of Bernoulli: Lets say there is a question with 4 multiple choice options.
Only one choice is correct.

- If you get the question correct it is 1 (4)
- If you get the question wrong, it is 0. (1,2,3)

So $X \sim Ber(p = 0.25)$

$$X(\Omega) = \{0,1\}, \quad \text{let } \Omega = \{1,2,3,4\}$$

$$P_X(x) = \begin{cases} 1/4 & \text{if } x = 1 \\ 3/4 & \text{if } x = 0 \end{cases}$$

$$P(X=0) = P(X^{-1}(0))$$
$$= P(\{1,2,3\}) = 3/4$$
$$P(X=1) = P(X^{-1}(1)) = P(\{4\}) = 1.$$

So $X : \Omega \to \mathbb{R}$

$$X(\omega) = \begin{cases} 1 & \text{if } \omega = 4 \\ 0 & \text{if } \omega = 1,2,3 \end{cases}$$

Defn 4.10 **Indicator random variable**:

The indicator random variable of an event A is the random variable $\mathbb{1}_A$ defined by

$$\boxed{\mathbb{1}_A = \begin{cases} 1 & \text{if } \omega \in A \\ 0 & \text{if } \omega \notin A \end{cases}}$$

Note: $P(\mathbb{1}_A = 1) = P(A)$ and $P(\mathbb{1}_A = 0) = P(A^c) = 1 - P(A)$ and so

$$\boxed{\mathbb{1}_A \sim \mathrm{Ber}(P(A))}$$

Defn 4.11: Binomial distribution:

We say that a random variable X has the binomial distribution, with parameters n and p and write

$$X \sim Bin(n,p)$$

if

$$X(\Omega) = \{0,1,2,3,\dots,n\}$$

and it has mass function

$$\boxed{p_X(k) = P(X=k) = \begin{cases} \binom{n}{k} p^k (1-p)^k & \text{if } k=0,1,\dots,n \\ 0 & \text{otherwise} \end{cases}}$$

Here $\binom{n}{k}$ is the binomial co-efficient also denoted C^n_k

Note if $n=1$, Binomial becomes $Ber(p)$.
if $n=1$ $\quad X \sim Bin(1,p)$

$$p_X(k) = P(X=k) = \begin{cases} \binom{1}{k} p^k (1-p)^k \\ 0 \end{cases}$$

only 2 probability choices hence $Ber(p)$.

Note: Binomial distribution is just repeating independant Bernoulli trials n times. For more detail, see chapter 9.

The fact is that the mass function for binomial distribution follows from the binomial theorem:

for each $n \in \mathbb{N}$, and $a, b \in \mathbb{R}$,

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k b^{n-k}$$

which gives

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

$$= \left[p + (1-p)^n\right] = 1$$

which shows binomial distribution satisfies (M1) and (M2).

Example 4.12: Suppose a coin has probability of p showing heads and is tossed n times.
The sample space is

$$\Omega = \{H,T\}^n = \{(w_1, \dots, w_n) \mid w_i \in \{H,T\}\}$$

i.e. Ω is the set of all n-tuples with entries taken from $\{H,T\}$. Associated to the toss, the ith toss is an indicator random variable.

$$X_i = \mathbb{1}_{i\text{th toss gives heads}}$$

$$= \mathbb{1}_{\{w_i = H\}} \sim \mathrm{Ber}(p)$$

Toss are <u>independant</u> (outcome of one toss does not affect the others) and the random variable

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

represents the total number of heads.

The event $\{X = k\}$ is the event of getting exactly k heads out of n tosses.

This set contains $\binom{n}{k}$ outcomes,

↳ counting, number of ways k heads can occur

[$\binom{n}{k}$ is vectors of length n containing k Heads and $n-k$ Tails), each of which has the same probability of occuring (because order of outcomes do not matter). Thus

$$P(X=k) = \binom{n}{k} P(\{(\underbrace{HH\dots H}_{k \text{ times}}\underbrace{TT\dots T}_{n-k \text{ times}})\})$$

$$= \binom{n}{k} \underbrace{pp\cdots p}_{k \text{ times}} \underbrace{(1-p)(1-p)\dots(1-p)}_{n-k \text{ times}}$$

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

i.e. $X \sim Bin(n,p)$

In general, the sum of independant $Ber(p)$-distributed random variables has the

$Bin(n,p)$ distribution.

Example: <u>Continuing multiple choice example:</u>

Say you attend a multiple choice exam. It consists of 10 multiple choice questions with 4 alternatives and one is correct.
You will pass the exam if you answer 6 or more questions correctly.
You decide to answer each of the question in a random way, in such a way that one question is not affected by other (<u>independant</u>).

What is the probability of a pass

Solution: Setting $i = 1, 2, \dots, 10,$

$$R_i = \begin{cases} 1 & \text{if } i^{th} \text{ answered correctly} \\ 0 & \text{if } i^{th} \text{ answered incorrectly.} \end{cases}$$

a question is an event

It can be seen that each question R_i is a Bernoulli distribution with the probability that

$$P(R_i = 1) = 1/4$$

$$P(R_i = 0) = 1 - 1/4 = 3/4$$

Let Random Variable X denote the number of correct answers.
It is given by

$$X = R_1 + R_2 + R_3 + R_4 + R_5 + R_6 + R_7 + R_8 + R_9 + R_{10}$$

$\Rightarrow$

$$X = \sum_{i=1}^{10} R_i$$

Clearly since R_i can attain a 1 or 0, X can attain values between 0 and 10.

- Consider first case $X = 0$:

Since answers do not influence each other, we conclude that events

$\{R_1 = a_1\}, \ldots, \{R_{10} = a_{10}\}$ is independant for every choice of the a_i is 0 or 1. (bernoulli events)

We find:

$$P(X=0) = P(\text{not a single } R_i \text{ (bernoulli event) equals } 1)$$
$$= P(R_1=0, R_2=0, \dots R_{10}=0)$$
$$= P(R_1=0) \cdot P(R_2=0) \dots P(R_{10}=0)$$
$$= \left(\frac{3}{4}\right)^{10}$$

which is the probability that answers for all questions are not correct.

The probability we answered exactly one question correctly equals

$$P(X=1) = \frac{1}{4} \cdot \left(\frac{3}{4}\right)^9 \cdot \boxed{10}$$

→ reason for 10 multiplies given below

which is the probability that the answer is correct times the probability that the other nine answers are wrong, times the <u>number of times this can occur</u>.

occurs 10 different ways

$$\left.\begin{aligned} P(X=1) = {} & P(R_1=1)\cdot P(R_2=0)\cdot P(R_3=0)\cdots P(R_{10}=0) \\ & + P(R_1=0)\cdot P(R_2=1)\cdot P(R_3=0)\cdots P(R_{10}=0) \\ & + P(R_1=0)\cdot P(R_2=0)\cdot P(R_3=1)\cdots P(R_{10}=0) \\ & \quad \vdots \\ & + P(R_1=0)\cdot P(R_2=0)\cdot P(R_3=0)\cdots P(R_{10}=1) \end{aligned}\right.$$

In general, we find that for $k=1,2,\cdots,10$, again using independance,

$$P(X=k) = \left(\frac{1}{4}\right)^k \cdot \left(\frac{3}{4}\right)^{10-k} \cdot \boxed{C^{10}_k} - \binom{10}{k}$$

10 choose k.

which is the probability that k questions were answered times 10-k answered wrong times the number of ways it can occur

$\hookrightarrow C^{10}_k$

More generally, we have to choose k different objects out of an ordered list of n objects, then:

- for the first n objects you have n possibilities, no matter which object you pick

- for the second, there are $n-1$ possibilities

- for the third there are $n-2$ and so on, ...

with $n-(k-1)$ for the k^{th}

So there are

$$\boxed{n(n-1)(n-2)\dots(n-k+1)}$$

ways to choose k objects

In how many ways can we choose 3 objects when the order matters?
There are $10 \cdot 9 \cdot 8$ ways

However order here does not matter:
To answer questions 2,5,8 correctly is the same as answering questions 8,5,2 correctly and so on.
The triplet $\{2,5,8\}$ can be chosen with $3.2.1$ different ways, all with same result.
(There are 6 permutations of 2,5,8)

Thus compensating for the 6 fold overcount, the number $C_{10,3}$ of ways to correctly answer 3 questions out of 10 becomes

$$C_3^{10} = \frac{10.9.8}{3.2.1}$$

More generally for $n \geq k$, and $1 \leq k \leq n$,

$$C_k^n = \frac{n!}{k!\,(n-k)!}$$

Defn 4.13: Geometric Distribution:

We say that a random variable X has a geometric distribution with parameter $p \in [0,1]$ and write

$$\boxed{X \sim \mathrm{Geo}(p)}$$

if

$$\boxed{X(\Omega) = \mathbb{N}}$$

has mass function:

$$\boxed{p_X(n) = \begin{cases} p(1-p)^{n-1} & \text{if } n \in \mathbb{N} \\ 0 & \text{otherwise} \end{cases}}$$

Let us determine the distribution function

$$F_X(x) = P(X \leq x)$$

$$= \sum_{\substack{n \in \mathbb{N} \\ n \leq x}} P(X=n) = \sum_{n=1}^{\lfloor x \rfloor} P(X=n)$$

Notation: In the last equality, the notation $\lfloor x \rfloor$ represents the largest integer smaller or equal to x.

Using the probability mass function given above, we then find for $x \geq 1$ that

$$F_X(x) = \sum_{n=1}^{\lfloor x \rfloor} p(1-p)^{n-1}$$

$$= p \sum_{n=1}^{\lfloor x \rfloor} (1-p)^{n-1}$$

substituting $m = n-1$ and changing index

$$= p \sum_{m=0}^{\lfloor x \rfloor - 1} (1-p)^{m}$$

We can now use formula for sum of a geometric progression

$$\sum_{m=0}^{N} a^m = \frac{1-a^{N+1}}{1-a} \quad \text{for } a \in [0,1)$$

Using geometric progression formula with $N = \lfloor x \rfloor - 1$, $a = 1-p$, this gives for $x \geq 1$,

$$F_X(x) = p\,\frac{1-(1-p)^{\lfloor x \rfloor}}{1-(1-p)} = 1-(1-p)^{\lfloor x \rfloor}$$

The distribution of x is thus given by

$$\boxed{F_X(x) = P(X \leq x) = \begin{cases} 0 & x < 1 \\ 1-(1-p)^{\lfloor x \rfloor} & x \geq 1 \end{cases}}$$

We see that this has the required property of distribution function that

$$\lim_{x\to\infty} F_X(x) = \lim_{x\to\infty} 1-(1-p)^{\lfloor x \rfloor}$$

$$= \lim_{x\to\infty} 1 - \underbrace{\lim_{x\to\infty} (1-p)^{\lfloor x \rfloor}}_{=0\ *}$$

$$= 1$$

* since $p \leq 1 \Rightarrow 1-p \leq 0$

The geometric distribution is the number of bernoulli trials to get one success.

Example: For example a bernoulli trial of tossing a coin. Let p be heads and $1-p$ be tails and number of trials it takes to get heads be n.

Then

$$p_X(n) = \underbrace{(1-p)(1-p)\ldots(1-p)}_{n-1 \text{ failures}} \cdot \underbrace{p}_{\text{success on } n^{th} \text{ trial}}$$

$$= (1-p)^n \, p$$

Let each bernoulli trial be

$$X_i = \begin{cases} p & \text{if } \omega = H \\ 1-p & \text{if } \omega = T \end{cases} \quad \text{for } i = 1, 2, \ldots, n.$$

For $p_X(n) = P(X=n)$

$$= P(\text{no of tails in } n-1 \text{ trials and heads on } n^{th} \text{ trial})$$

$$= P(\{X_1 = T\}\{X_2 = T\} \ldots \{X_{n-1} = T\}\{X_n = T\})$$

$$= P(X_1 = T) \cdot P(X_2 = T) \dots P(X_n = H)$$

due to independance

$$= (1-p)^n \cdot p$$

Example 4.1: You flip a biased coin with $P(\{H\}) = p$ until you get the first Heads.

Let X be the number of the flip on which you get your first head. Then

$$P(X = n) = P(\{T, T, T, \dots, T, H\})$$
$$= (1-p)^{n-1} p$$

Thus $X \sim Geo(p)$

This example illustrates that a geometric distribution describes the waiting time until a success in a series of Bernoulli trials

The geometric distribution has the **memoryless property**:

Theorem 4.15: Let X be a random variable that has geometric distribution $X \sim Geo(p)$. Then for any $n, k \in \mathbb{N}$,

$$P(X > n+k \mid X > k) = P(X > n)$$

proof: $P(X > n+k \mid X > k)$

by defn 3.2 $$= \frac{P(\{X > n+k\} \cap \{X > k\})}{P(X > k)}$$

because $\{X > n+k\} \subset \{X > k\}$ $$= \frac{P(X > n+k)}{P(X > k)}$$

using the distribution function given $$= \frac{(1-p)^{n+k}}{(1-p)^k}$$

$$= (1-p)^n = P(X > n)$$

$\blacksquare$

Explanation of memoryless property:

In practise, this memoryless property means that for how many trials you have already waited does not affect how many trials you will have to wait.
If you have had bad luck for a long time in a game, and had to wait to wait for that 6 from the die for a long time, that does not have the effect that the 6 is now bound to come soon.

To make it specific, let k be 5, and $n=10$. If our probability distribution is memoryless, the probability of $X>15$ if we know $X>5$ is the exact same as $X>10$.

Meaning that no matter how many trials you have had (k), it will not affect the number of trials you need in future (n).

Defn 4.16 : **Poisson Distribution:**

We say that the random variable X has the Poisson Distribution with parameter $\lambda > 0$ and write

$$\boxed{X \sim \mathrm{Pois}(\lambda)}$$

if

$$X(\Omega) = \{0, 1, 2, \dots\}$$

and it has mass function

$$\boxed{p_X(n) = \begin{cases} \frac{\lambda^n}{n!} e^{-\lambda} & \text{if } n = 0, 1, 2, \dots \\ 0 & \text{otherwise.} \end{cases}}$$

The fact that above is indeed a mass function follows from the taylor series for the exponential function for each $x \in \mathbb{R}$

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

Thus,

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

$$= e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^n}{n!}$$

$$= e^{-\lambda} e^{\lambda} = 1$$

Thus satisfying (m1) and (m2).