ST120

Haria

October 2025

1 Lectures 1 + 2

1.1 Probability Spaces

Consider a simple probability problem. I have a bag with 20 balls of three different colours 4 red balls 6 blue balls and 10 green balls. A simple question to be asked is what is $\mathbb{P}(\text{red})$. From intuition you know this number is $\frac{4}{20}$ but it is important to formalise what's going on.

1.1.1 Sample Space

We can say that red green and blue are three **outcomes** that can occur. It can be useful to collect these three events together into a set as such $\{\text{red}, \text{blue}, \text{green}\}$.

Definition 1. *Let Ω denote the Sample Space, The set of all outcomes.*

1.1.2 Event Space

Sometimes we want to discuss what we call and even instead of a specific outcome. Sometimes we may want to ask what is $\mathbb{P}(\neg\text{red})$ we can associate $\neg$red with the set $\{\text{blue}, \text{green}\}$. We call such a thing event and events are sets of outcomes and as such every event is a subset of the Sample Space

Definition 2. *Let $\mathcal{F}$ denote the Event Space. The Set of all events. $\mathcal{F} \subseteq \mathcal{P}(\Omega)$*

1.1.3 Probability Map

The last thing we want to do is be bale to go from an event to the actual probability.

Definition 3. *Let $\mathbb{P}$ denote the probability map. It is a function $\mathcal{F} \to [0,1]$ associating an event with a probability*

Such a map should fulfil some list of properties as follows

- $\mathbb{P}(\emptyset) = 0$
- $\mathbb{P}(\Omega) = 1$

- For any events $A, B \in \mathcal{F}$ where $A \cap B = \emptyset$ then $\mathbb{P}(A) + \mathbb{P}(B) = \mathbb{P}(A \cup B)$

Definition 4. *A Probability Space is the triple* $(\Omega, \mathcal{F}, \mathbb{P})$

1.2 Uniform Probability Spaces

Definition 5. *A uniform probability space is a triple* $(\Omega, \mathcal{F}, \mathbb{P})$ *satisfying the normal properties aswell as a uniformity where for any* $\omega, \omega' \in \Omega$ *we have* $\mathbb{P}(\omega) = \mathbb{P}(\omega')$

As such we conject that $\mathbb{P}(A) \propto |A|$

Proposition 1. *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *denote a UPS then for all* $\omega \in \Omega$ $\mathbb{P}(\{\omega\}) = \frac{1}{|\Omega|}$

Proof. Since for all $\omega, \omega' \in \Omega$ we have $\mathbb{P}(\omega) = \mathbb{P}(\omega')$ let $p = \mathbb{P}(\omega)$

$$1 = \mathbb{P}(\Omega) = \sum_{\omega \in \Omega} \mathbb{P}(\omega) \ = \sum_{\omega \in \Omega} p = p|\Omega|$$

$$p = \frac{1}{|\Omega|}$$

□

Proposition 2. *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *denote a UPS then for all* $A \in \mathcal{F}$ $\mathbb{P}(A) = \frac{|A|}{|\Omega|}$

Proof.

$$\mathbb{P}(A) = \sum_{\omega \in A} \mathbb{P}(\omega) = p|A| = \frac{|A|}{|\Omega|}$$

□

1.3 Basics of Combinatorics

Here we shall present the Fundamental rules of counting

- Correspondence Rule : If A and B are in 1 - 1 correspondence then $|A| = |B|$
- Addition Rule : if $A_1, A_2, \ldots, A_n$ are pairwise disjoint then

$$\left|\bigcup_{i=1}^{n} A_i\right| = \sum_{i=1}^{n} |A_i|$$

- Fundamental Counting Principle : Suppose a finite set E can have its elements determined in k successive steps with n_i possibilities for each step $1, \ldots, k$ and different choices lead to different elements then $|E| = \prod_{i=1}^{k} n_i$

2 Lecture 3 - Birthdays Problems

We have started to build up a theory for how to calculate probabilities. Now we will try to apply these and also see whatever definitions are needed to fill in the gaps

2.1 Tuples and Orderings

2.1.1 Tuples

The first problem we will attempt to solve is "What is the probability, given a room of 30 people, that at least two of them share a birthday/ The first thing we will want to identify is what is the sample space. The sample space intuitively should be the space of all configurations of birthdays the group can have. Naturally this forms a uniform probability space as each outcome is equally likely. We also have to consider how to mathematically represent these outcomes so that we can find the size of the entire sample space. The way we shall represent these are as 30-tuples $(b_1, b_2, \dots, b_{30})$ where b_i is the ith persons birthday.

Now lets discuss what a tuple is

Definition 6. *Given a set A such that $|A| = n \in \mathbb{N}$. A sequence of length $k \in \mathbb{N}$ of elements of A is an ordered k-tuple $(a_1, \dots a_k)$ such that $a_i \in A$ where $i \in \{1, \dots, k\}$ We denote these sequences $S_{n,k}(A)$*

Now we need to be able to compute the cardinality of sets of sequences

Proposition 3. *For a set A where $|A| = n \in \mathbb{N}$*

$$|S_{n,k}(A)| = n^k$$

Proof. To construct an arbitrary element of $S_{n,k}(A)$ we first choose a_1 for which we have n options then we choose a_2 for which we have n options and so on up till n_k where we have n options. As each option provides a different element in the final set we can use the fundamental principal of counting to compute the size.

$$|S_{n,k}(A)| = \underbrace{n \times n \times \cdots \times n}_{k \text{ times}} = n^k$$

$\square$

Using this result we can find the size of the sample space in the birthdays problem as $|\Omega| = 365^{30}$

2.1.2 Orderings

Next we wish to compute the cardinality of the event *"At least two people share a birthday"*. Much easier than computing this is to compute the complementary event *"No two people share a birthday"*. this means that each tuple in this event has no repeated elements. this is what we call and ordering.

Definition 7. *Given a set A such that $|A| = n \in \mathbb{N}$ and a $k \leq n$ we can say and ordering of length k of elements of A is a sequence of length k of a with no repetition. We denote these $O_{n,k}(A)$*

$$O_{n,k}(A) = (a_1, \ldots, a_k) : a_i \in A, \forall_{i,j}\ a_i \neq a_j$$

Now we would also like to be able to compute the sizes of sets of orderings.

Proposition 4. *Given a set A such that $|A| = n \in \mathbb{N}$ and a k such that $k \leq n$*

$$|O_{n,k}(A)| = \frac{n!}{(n-k)!}$$

Proof. We can determine elements of $O_{n,k}$ elements by element. To determine a_1 we have n choices. For the a_2 we have $n-1$ choices as we cant repeat the first element. For a_k we have $n-(k+1)$ choices. As each series of choices leads to a different element we can apply the fundamental principle of counting.

$$|O_{n,k}(A)| = n \times (n-1) \times \cdots \times (n-k+1)_{=} \frac{n!}{(n-k)!}$$

□

So if we let B denote the event no one shares a birthday then $B = O_{365,30}(365 \text{ days})$ so $|B| = 365 \times \cdots \times (365 - 30 + 1)$ so

$$\mathbb{P}(B) = \frac{|B|}{|\Omega|} \approx 0.29$$

So we then have for our original event that at least 2 people share a probability of about 0.71.

2.2 Another Example

Now lets consider a similar problem of there being **exactly** 2 people sharing a birthday. We can characterize each outcome in our desired event by three characteristics

1. The two people who share a birthday
2. The day that they share
3. The days everyone else has

Now to compute the size of the event by the fundamental principle of counting we need to compute the number of choices for each piece of information.

1. An intuitive guess for the first one is $|O_{30,2}|$ but this cares about the order that the two people share a birthday are picked in which doesn't really matter so we can divide this by the ways to order the two yielding us

$$\frac{|O_{30,2}|}{|O_{2,2}|} = \frac{30!}{28!2!}$$

2. There are 365 possibilities for this day
3. For the remaining 28 people by the Fundamental Counting Principle we have $364 \times \cdots \times (365-28)$

Now via computation and the same sample space as before we get the probability is roughly 0.28

2.3 Combinations

Definition 8. *Let A be a set such that $|A| = n \in \mathbb{N}$. We can say that a combination of k elements of A is a subset of A with k elements. Let the set of combinations be denoted $C_{n,k}(A)$*

Proposition 5. *Let A be a set such that $|A| = n \in \mathbb{N}$ and $k \leq n$*

$$|C_{n,k}(A)| = \binom{n}{k}$$

Proof. On ordering of length k can be obtained uniquely in the following steps.

1. Choose from $C_{n,k}(A)$. This gives $|C_{n,k}(A)|$ choices
2. Choose a permutation of these. $k!$ choices.

By the FPC we get

$$|O_{n,k}(A)| = k!|C_{n,k}(A)|$$

By rearranging we can get

$$|C_{n,k}(A)| = \frac{n!}{(n-k)!k!} = \binom{n}{k}$$

□

As an exemplar problem we can consider all the ways that, having rolled 8 fair dice, our outcome has three twos, three fours and two fives. We can note that each of our valid outcomes is uniquely characterised by what dice gives us the twos and fours. Initially we have $\binom{8}{3}$ positions for the twos then $\binom{5}{3}$ for the fours and the fives must go where is left. By the FPC this event has size being the product of these two numbers

3 Lecture 4 - Partitions

Definition 9. *Let A be a set such that $|A| = n \in \mathbb{N}$ and let $r \leq n$.*
An ordered parttion of A into r ordered subsets of cardinalt $k_1, \ldots, k_r$ is a sequence of $(A_1, \ldots, A_r)$ such that $|A_i| = k_i$, all the elements are pairwise disjoint and $\bigcup_{i=1}^{r} A_i = A$

Proposition 6. *LEt A be a set such that $|A| = n \in \mathbb{N}$ and $r \leq n$. The number of partitions of A into r subsets of cardinality $k_1, \ldots, k_r$ is*

$$\frac{n!}{k_1! \times \cdots k_r!}$$

Proof. We can determine a partition as follows
Choose elements as follows

$$A_1 \subseteq A \text{ such that} |A_1| = k_1 \text{ gives } \binom{n}{k_1}$$
$$A_1 \subseteq A \backslash A_1 \text{ such that} |A_1| = k_1 \text{ gives } \binom{n - k_1}{k_2}$$
$$\ldots$$
$$A_1 \subseteq A \backslash \bigcup_{i=1}^{r-1} A_i \text{ such that} |A_1| = k_1 \text{ gives } \binom{n - \sum_{i=1}^{r-1} k}{k_r}$$

By application of the fundamental priniple of counting and some carful counting we get the amount of partitions is

$$\frac{n!}{k_1! \times \cdots k_r!}$$

□

3.1 Samplings

Sampling is how we produce realisations of random variables.

3.1.1 Example 1

Imagine a population of size $n \in \mathbb{N}$ and there are n_1 type 1 individuals and n_2 type two individuals.
Suppose we draw a sample of size $k \leq n$ from the population without replacement. The probability of a sample containing k_1 type one individuals and $k_2 = k - k_1$ of type 2 is

4 Lectrue 5 - General Probability Spaces

We have spent a while specifically discussing uniform probability spaces. There are some more gnerealisations we may want to make to how our space works

- Let Ω be an arbitrary set

- Allow $\mathcal{F}$ to reflect partial knowledge
- Following from the previous point we want to appreciate that $\mathcal{F} = \mathcal{P}(\Omega)$ is not always feasible.
- Finally we want to define porperties of $\mathbb{P}$ on general $\mathcal{F}$

4.1 Sample and Event Spaces

Definition 10. *Ω is the set of all possible outcomes for a propabalistic process.*

In this course three cases are covered

- $|\Omega| = n \in \mathbb{N}$
- $|\Omega| = |\mathbb{N}|$
- $|\Omega| = |\mathbb{R}|$

  Definition 11. *$\mathcal{F}$ is the event space and contains collections of outcomes for which you can make a judgement on wether the event happened or not*

Example 1. *Imagine roilling a dice $\Omega = \{1, 2, 3, 4, 5, 6\}$. The only information you are told is whether or not the outcome is odd or even. As this is the only information you are given you cannot diffrentiate the events $\{2\}$ and $\{4\}$. As such they cannot be in the event space. Here $\mathcal{F} = \{\emptyset, \{1, 3, 5\}, \{2, 4, 6\}\}$.*

Now lets present a more formal definition of the event space

Definition 12. *Let Ω be a sample scpae and $\mathcal{F} \subseteq \mathcal{P}(\Omega)$*
We can say $\mathcal{F}$ is an event space iff

1. $\Omega \in \mathcal{F}$, $\emptyset \in \mathcal{F}$
2. *If $A \subseteq \Omega$ and $A \in \mathcal{F}$ then $A^C \in \mathcal{F}$ (closure under complements)*
3. *Given a set $A = \{A_n \in \mathcal{F} | n \in \mathbb{N}\}$ then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$ (closure ubder coubntable union)*

Proposition 7. *$\Omega \neq \emptyset$ and $\mathcal{F} = \mathcal{P}(\Omega)$ then $\mathcal{F}$ is an event space*

Proof.
1. $\Omega \in \Omega \Rightarrow \Omega \in P(\Omega) = \mathcal{F}$
2. $A \subseteq \Omega, A \in \mathcal{F}$ then $A^C = \Omega \backslash A \subseteq \Omega \Rightarrow A^C \in \mathcal{P}(\Omega) = \mathcal{F}$
3. Consider $A = \{A_n \in \mathcal{F} | n \in \mathbb{N}\}$

$$\bigcup_{A_n \in A} A_n \subseteq \bigcup_{A_n \in A} \Omega = \Omega \in \mathcal{P}(\Omega) = \mathcal{F}$$

□

We can now also look at some derived properties of the event space inhering from these defined properites

Proposition 8. *LEt $\mathcal{F}$ be an event space then*

1. *$\mathcal{F}$ is closed under finite union*
2. *$\mathcal{F}$ is closed under finite intersection*
3. *$\mathcal{F}$ is closed under countable intersection*

Proof.

1. Let $A_1, \dots, A_n \in \mathcal{F}$. Set $A_j = \emptyset, j > n$. therefor For all $i \in \mathbb{N}^+, A_i \in \mathcal{F}$

$$\bigcup_{j=1}^{n} A_j = \bigcup_{j=1}^{\infty} A_j \in \mathcal{F}$$

2. LEt $A_1, \dots, A_n \in \mathcal{F}$. By demorgans law we may state

$$\bigcap_{j=1}^{n} A_j = \left(\bigcup_{j=1}^{n} A^C \right)^C$$

By the closure of $\mathcal{F}$ under complements and under finite union we can say this finite intersection is in $\mathcal{F}$

3. For the infinite case we apply the same argument we just used only swapping finite for infinite

□

5 Lecture 6 - Probability measure

5.1 Probability measure

Definition 13. *Given samples and event spaces $\Omega, \mathcal{F}$ a function $\mathbb{P} : \mathcal{F} \to \mathbb{R}$ is a probability measure if*

1. *For all $x \in \mathcal{F}, \mathbb{P}(x) \in [0, 1]$*
2. *$\mathbb{P}(\Omega) = 1$*
3. *For all m, n such that $m \neq n, A_n \cap A_m = \emptyset$ then*

$$\mathbb{P}\left(\bigcup_{j=1}^{\infty} A_j \right) = \sum_{j=1}^{\infty} \mathbb{P}(A_j)$$

Proposition 9. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. We can make then the following propoistions about $\mathbb{P}$*

1. $\mathbb{P}(\emptyset) = 0$
2. *if* $A, B \in \mathcal{F}$ *and* $A \subseteq B$ *then* $\mathbb{P}(B - A) = \mathbb{P}(B) - \mathbb{P}(A)$
3. *For all* $A \in \mathcal{F}, \mathbb{P}(A^C) = 1 - \mathbb{P}(A)$

Proof. 1. For all $n \geq 1$ let $A_n = \emptyset$. Therefore we can state $\bigcup_{n=1}^{\infty} A_n = \emptyset$

$$\mathbb{P}(\emptyset) = \mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}A_n = \sum_{n=1}^{\infty} \mathbb{P}(\emptyset)$$

This only holds if $\mathbb{P}(\emptyset) = 0$

2. Using the same way we proved closure of finite unions in the event space we may prove a finite additivity for disjpin sets. This fact will be assumed here. Let $B = (B - A) \cup B$. By the definition of set difference we know that $(B - A) \cap A = \emptyset$. This allows us to apply finite additivity

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

3. $\mathbb{P}(A^C) = \mathbb{P}(\Omega - A) = \mathbb{P}(\Omega) - \mathbb{P}(A) = 1 - \mathbb{P}(A)$

□

5.2 Principle of inclusion Exclusion

Proposition 10. *The Principle of Inclusion Exclusion is as follows. Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space. Then for* $A_1, \ldots, A_n \in \mathcal{F}$

$$\mathbb{P}\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{k=1}^{n} \left((-1)^{k+1} \sum_{2 \leq i_1 < \cdots < i_k \leq n} \mathbb{P}(A_{i_1} \cap \cdots \cap a_{i_k}) \right)$$

Proof. A proof sketchfor $n = 2$ is given here

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

By application of finite additivity PIE for $n = 2$ may be gotten. □

Proposition 11. *If* $(\Omega, \mathcal{F}, \mathbb{P})$ *is a propbabilty space then if* $A, B \in \mathcal{F}$ *and* $A \subseteq B$ *Then* $\mathbb{P}(A) \leq \mathbb{P}(B)$

Proposition 12. *Booles inequality : If* $(\Omega, \mathcal{F}, \mathbb{P})$ *is a propbabilty space then for a list* $A_1, \ldots, A_n \in \mathcal{F}$ *then*

$$\mathbb{P}\left(\bigcup_{k=1}^{n} A_n\right) \leq \sum_{k=1}^{n} \mathbb{P}(A_n)$$

Proof. By induction on n

- Base case $n = 2$. $\mathbb{P}(A_1 \cup A_2) = \mathbb{P}(A_1) + \mathbb{P}(A_2) - \mathbb{P}(A_1 \cap B_2) \leq \mathbb{P}(A_1) + \mathbb{P}(A_2)$

- Assume its true for some $n = k$.

$$\begin{aligned}\mathbb{P}(\bigcup_{j=1}^{k+1} A_j) &= \mathbb{P}(\bigcup_{j=1}^{k} A_j \cup A_{k+1})\\ &= \mathbb{P}(\bigcup_{j=1}^{k} A_j) + \mathbb{P}(A_{k+1}) - \mathbb{P}(\bigcup_{j=1}^{k} A_j \cap A_{k+1})\\ &\leq \mathbb{P}(\bigcup_{j=1}^{k} A_j) + \mathbb{P}(A_k + 1)\\ &\leq \sum_{j=1}^{k+1} \mathbb{P}(A_j)\end{aligned}$$

□

6 Lecture 7 - Conditional Probability

Suppose we are doing an experiment and an we have knowledge that an event B has happened. Knowing this how can we find the probability that another event A has happened. We call theis the prbability of A given B with this event denoted as follows

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

Proposition 13. *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $B \in \mathcal{F}$ Define $\mathbb{P}_B : \mathcal{F} \to \mathbb{R}$ as $\mathbb{P}_B(A) = \mathbb{P}(A|B)$. $\mathbb{P}_B$ is a probability measure*

Proof. There are three propterites to be checked.

1. It is known that $\mathbb{P}(A) \in [0,1]$ for all $A \in \mathcal{F}$. We also know that for this A, $A \cap B \in \mathcal{F}$. This means that $\mathbb{P}(A \cap B)$ is defined. We know that $\mathbb{P}(A \cap B) \leq \mathbb{P}(B)$ Therefore
$$\frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \in [0,1]$$
2. $$\mathbb{P}_B(\Omega) = \frac{\mathbb{P}(B \cap \Omega)}{\mathbb{P}(B)} = 1$$
3. The next property to check is coiuntable additivity. Let for all $n, A_n \in \mathcal{F}$

and let them all be pariise disjoint.

$$\begin{aligned}\mathbb{P}_B(\bigcup_{n=1}^{\infty} A_n) &= \frac{\mathbb{P}_B(\bigcup_{n=1}^{\infty} A_n \cap B)}{\mathbb{P}(B)} \\ &= \frac{1}{\mathbb{P}(B)}\mathbb{P}(\bigcup_{n=1}^{\infty} A_n \cup B) \\ &= \sum_{n=1}^{\infty} \frac{\mathbb{P}(A_n \cup B)}{\mathbb{P}(B)} \\ &= \sum_{n=1}^{\infty} \mathbb{P}_B(A_n)\end{aligned}$$

This assumed we were able to aplpy countable additivity to the orginal measure which requires $A_n \cap B$ to be pariwise disjoint

$$(A_n \cap B) \cap (A_m \cap B) \subseteq A_n \cap A_m = \emptyset$$

□

This equation leads us to the multiplication rule

Theorem 1. *Multiplication rule :* $\mathbb{P}(A_1 \cap \cdots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_1|A_2)\mathbb{P}(A_3|A_1 \cap A_2)\cdots$

this is proven using an expansion of the law of consitionak probability

7 Lectire 8 + 9 : Firther applications

If we wanted to know the probability of an event C happening. We can consider all of the disjoint cases in which it could happen and take some kind of weighted mean of all of these probabilities to find C. This is the law of toal probability.

Theorem 2. $\mathbb{P}(C_2) = \mathbb{P}(C_2|C_1)\mathbb{P}(C_1) + \mathbb{P}(C_2|C_1^C)\mathbb{P}(C_1^C)$

Proof. Diue to the fact $(C_1 \cup C_2) = \Omega$ The event $C_2 = (C_2 \cap C_1) \cup (C_2 \cap C_1^C)$. Then by application of finite additivity and the law of conditoinal probability the law of total probability can be derived □

This result can be futher generalised

Theorem 3. *Law of Total probability: Let* $\{B_n|n \in [[N]]\}$ *be a partition of* Ω*. We can say*

$$\mathbb{P}(A) = \sum_{n=1}^{N} \mathbb{P}(A|B_n)\mathbb{P}(B_n)$$

The proof is done in the same way by the aprreciation of the union of $B_n = \Omega$ and then applying countable additivitty and the laws of consitional probability.

8 Lecture 10 - Random Variables

Consider having done a random experiment which is quite complex as is the sample scpase (Ω) as such the indiviual events ($\mathcal{F}$) are also complex but these events are such that we know wether they have happened or not.

Very often we wish to abstract these elements of Ω to something much simpler i.e something that has a numerical value given by the random process.

We can express this random variable as a function $X : \Omega \to \mathbb{R}$. This course will only consider random variables into the real numbers.

Definition 14. *Random Variable: Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A real-valued random variable is a function $X : \Omega \to \mathbb{R}$ such that $\{\omega \in \Omega | X(\omega) \leq a\} \in \mathcal{F}$ for all $a \in \mathbb{R}$ (This says that specifying conditions via X gives an event that we can determine wether it has happened or not)*

Example 2. *Consider rolling tow dice. You win money given by given by the greater value amongst the dice rolled. In this case*

$$\Omega = \{1, 2, 3, 4, 5, 6\}^2$$

$$X(x, y) = max(x, y)$$

It can be seen that using the random varable X simplifies the outcomes in Ω

$$\mathbb{P}(X = \overline{y}) = \mathbb{P}(\omega \in \Omega | X(\omega) = 5) = \frac{1}{5}$$

We can see at the end we use a condition on X in order to specify an event. Intead in an inequality an equality was used but this can be shown to be inside of $\mathcal{F}$ by closere of $\mathcal{F}$ under set difference

We can consider the a new probability measure on $\mathbb{R}$ induced by the random variable X and if this is all we care about we can leave the old prbability space behond and only cosider $\mathbb{R}$ as our sample space

Definition 15. *Distrobution: The distrobution of a random variable X is the probability measure on $\mathbb{R}$ given by*

$$\mathbb{P}_X(B) = \mathbb{P}(\{\omega \in \Omega | X(\omega) \in B\}$$

such that $B \in \mathbb{R}$

Example 3. *Cont.* $\mathbb{P}_X(\{k\}) = \frac{2k-1}{36}$ *for* $k = 1, 2, 3, 4, 5, 6$
otherwise $\mathbb{P}(\mathbb{R} \setminus \{1, 2, 3, 4, 5, 6\}) = 0$

Definition 16. *Event Space: We will usually denote the envent space on $\mathbb{R}$ denoted $\mathcal{B}(\mathbb{R})$ contains singletons $\{x\}$ and intervals $(a, x]$ And is close under countable set operations as is $\mathcal{F}$. We have no need to be able to lost elements.*

Definition 17. *The triplet $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_X)$ is a simpler derived probability space.*

Proposition 14. *The function $\mathbb{P}_X$ is a proabbility measure on R*

Proof. We must check the three conditions of probability measure

1. $\mathbb{P}_X(B) = \mathbb{P}(\{\omega \in \Omega | X(\omega) \in B\}) \in [0,1]$
2. $\mathbb{P}_X(\mathbb{R}) = \mathbb{P}(\{\omega \in \Omega | X(\omega) \in \mathbb{R}\}) = \mathbb{P}(\Omega) = 1$
3. Suppose $B_1, B_2, \dots \in \mathcal{B}(\mathcal{R})$ are disjoint.

$$\begin{aligned}\mathbb{P}_X\left(\bigcup_{n=1}^{\infty} B_n\right) &= \mathbb{P}(\{\omega \in \Omega | X(\omega) \in \bigcup_{n=1}^{\infty} B_n\}) \\ &= \mathbb{P}(\bigcup_{n=1}^{\infty}\{\omega \in \Omega | X(\omega) \in B_n\}) \\ &= \sum_{n=1}^{\infty} \mathbb{P}(\{\omega \in \Omega | X(\omega) \in B_n\}) \\ &= \sum_{n=1}^{\infty} \mathbb{P}_X(B_n)\end{aligned}$$

This uses thhe facts that pre-image of the union is the union of the pre-image; that pre-image of disjoint sets are disjoint,

□

9 Lecture 11

Definition 18. *Let $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_X)$ be an induced probability space. We may it is discrete if there exist a countable or finite set $S \subseteq R$ such that $\mathbb{P}(X \in \mathbb{R} \setminus S) = 0$*

Definition 19. *Let $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_X)$ be a probility spacee and X be drv we define the prbability mass function $p_X : \mathbb{R} \to [0,1]$ as follows*

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

This is a simpler function as it allwos us to investigate individual probabilities instead of those fo a set.

Definition 20. *Let $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_X)$ be a probability space and X be a drv. We may define the discrete support of X or $\mathbb{P}_X$ as follows*

$$\{x \in \mathbb{R} | p_X(x) > 0\}$$

Proposition 15. *Let X be a drv and D_X be its support*

$$\mathbb{P}_X(B) = \sum_{x \in B \cap D^C} p_X(x)$$

Proof. We know there is a $S \subseteq \mathbb{R}$ such that $\mathbb{P}_X(S) = 0$. From here we can determine that D^C is countable as $D \subseteq S$. We may say that

$$P_X(B) = P_X(B \cap D) + P_X(B \cap D^C)$$

The second term has to be zero as $B \cap D^C \subseteq D^C$ so $0 \leq P_X(B \cap D^C) \leq P_X(D^C) \leq 0$
We can then evaluate as follows

$$\begin{aligned} P_X(B) &= P_X(B \cap D) \\ &= P_X\left(\bigcup_{x \in B \cap D} \{x\}\right) \\ &= \sum_{x \in B \cap D} P_X(\{x\}) \\ &= \sum_{x \in B \cap D} p_X(x) \end{aligned}$$

□

From here we have the fact that p_X caontains all the information of $\mathbb{P}_X$ as the latter can be reconstructed from the fomrer.

10 Lecture 12

As said before D_X doest necessarilly equal S

Example 4. *Consider rolling a die.* $\Omega = \{1, 2, 3, 4, 5, 6\}$, $X : \Omega \to \mathbb{R}, X(k) = k$. *From here we get the definition of* $D_X = \{1, 2, 3, 4, 5, 6\}$ *which is a suitable candidate for* S. *But this is not the only choice. Any countable extention of* D_X *is also a candidate for example* $S = \mathbb{N}$ *as there are no elements in* $\mathbb{N}^C$ *that are in* D_X *so* $\mathbb{P}_X(\mathbb{N}^C) = 0$

Definition 21. *A function* $f : \mathbb{R} \to [0, 1]$ *is a function is the set D defined as follows*

$$D = \{\{x | f(x) > 0\}\}$$

Is countable and $\sum_{k \in D} f(k) = 1$

This gives us a definition of a probability mass function without already having a space. We can however construct a space from here.

Proposition 16. *LEt* f *be a proabbaility mass function. Then there exist a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *and a drv* X *such that* $p_X = f$

Proof. Take D to be the discrete support of f. Let $\Omega = \mathbb{R}, \mathcal{F} = \mathcal{B}(\mathbb{R}), \mathbb{P}(B) = \sum_{x \in B \cap D} f(x)$
We now need to prove that $\mathbb{P}$ is a measure

- $$\mathbb{P}(\mathbb{R}) = \sum_{k \in \mathbb{R} \cap D} f(k) = \sum_{x \in D} f(x) = 1$$

- $\mathbb{P} \in [0,1]$
- Suppose $A_1, \ldots, A_n \in \mathcal{F}$ are all pairwise disjoint.

$$\begin{aligned}
\mathbb{P}\left(\bigcup_{n=1}^{\infty} A_n\right) &= \sum_{k \in \left(\bigcup_{n=1}^{\infty} A_n\right) \cap D} f(x) \\
&= \sum_{k \in \left(\bigcup_{n=1}^{\infty} A_n \cap D\right)} f(x) \\
&= \sum_{n=1}^{\infty} \sum_{x \in A_n \cap D} f(x) \\
&= \sum_{n=1}^{\infty} \mathbb{P}(A_n)
\end{aligned}$$

Now we need to construct the DRV. LEt $X(x) = x$. X has the right domain and codomain. and we can also say $\mathbb{P} = \mathbb{P}_X$. We also need to check it has an opporopriate countable set S For this we may try D.

$$\mathbb{P}_X(\mathbb{R} \setminus D) = \mathbb{P}_X(D^C) = \mathbb{P}(D^C) = \sum_{x \in D^C \cap D} f(x) = 0$$

Finally we need to check f is the probabilit mass function for X. If $x \in D$

$$p_X(x) = \mathbb{P}_X(\{x\}) = \mathbb{P}(\{x\}) = \sum_{x \in \{x\} \cap D} f(x) = f(x)$$

Otherwise

$$p_X(x) = \mathbb{P}_X(\{x\}) = \mathbb{P}(\{x\}) = \sum_{x \in \{x\} \cap D} f(x) = \sum_{x \in \emptyset} f(x) = 0 = f(x)$$

□

Now lets look at some experiments that are discrete spaces.

Example 5. *Imagine tossing a p-coin which is a coin where the probability of heads is p. We may construct our space as follows.* $\Omega = \{H, T\}, \mathbb{P}(\{H\}) = p, \mathbb{P}(\{T\}) = 1 - p$. *We may define* $X : \Omega \to \mathbb{R}$ *as* $X(H) = 1, X(T) = 0$. *We may then say* $X \sim$ *Bernoulli*(p)

Definition 22. *Where* $p \in [0,1]$ *we may say* $X \sim$ *Bernoulli*(p) *if*

$$p_X(x) = \begin{cases} p & x = 1 \\ 1 - p & x = 0 \\ 0 & \textit{otherwise} \end{cases}$$

Example 6. *Consider another experiment where we toss the p-coin till we gewt heads. Let X map to the number of tosses and* Ω *be the set of sequences of tails which then end in heads.*

$$\mathbb{P}(\{\underbrace{T\dots T}_{k\ times}H\}) = (1-p)^k p$$

$$X(\{\underbrace{T\dots T}_{k\ times}H\}) = k+1$$

Definition 23. *We can say* $X \sim Geom(p)$ *if*

$$p_X(k) = \begin{cases} p(1-p)^k - 1 & k \in \mathbb{N} \\ 0 & Otherwise \end{cases}$$

Definition 24. $X \sim Bin(n,p)$ *where* $n \in \mathbb{N}, p \in [0,1]$ *if*

$$p_X(k) = \begin{cases} \binom{n}{x} p^x (p-1)^{(n-x)} & k \in \{0,\dots,n\} \\ 0 & Otherwise \end{cases}$$

11 Lecture 13

Expectation: Sometimes we want to consider question such as what is the average outcome of a roll of a dice. in order to do this we may calculuate some form of a weighted mean.

$$\frac{1}{6}(1+2+3+4+5+6) = 3.5$$

We call this number the expectation of reolling a dice

Definition 25. *Consider a discrete random variable X with a suppoer D and a mass function* p_X *We define the expectation of X denoted* $\mathbb{E}(X)$ *defined as follows*

$$\mathbb{E}(X) = \sum_{x \in D} x p_X(x)$$

As long as it converges absolutley (sum of absolute values converges), Otherwise the expectation is undefined

Definition 26. *A drv X with support D and probability mass function* p_X *is integrable if the sum*

$$\sum_{x \in D} |x| p_X(x)$$

converges

Example 7. *Consider $X \sim B(n,p)$*

$$\begin{aligned}
\mathbb{E}(X) &= \sum_{k=0}^{n} k\binom{n}{k}p^k(1-p)^{n-k} \\
&= \sum_{k=1}^{n} n\binom{n-1}{k-1}p^k(1-p)^{n-k} \\
&= n\sum_{j=0}^{n-1}\binom{n-1}{j}p^{j+1}(1-p)^{n-j-1} \\
&= np\sum_{j=0}^{n-1}\binom{n-1}{j}p^{j}(1-p)^{n-j-1} \\
&= np(p+(1-p))^{n-1} \\
&= np
\end{aligned}$$

Comsider now fixing this value $\lambda = np$. This leads us to the following experiment. We may consider tossing the p-coin n times such that the expected number of heads is fixed.

$$\mathbb{P}(X=k) = \binom{n}{k}p^k(1-p)^{n-k} = \frac{n!}{k!(n-k)!}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k} = \frac{\lambda^k}{k!}\frac{n!}{n^k(n-k)!}\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^{-k}$$

Taking the limit of $n \to \infty$ we get this is equal to

$$\frac{\lambda^k}{k!}e^{-\lambda}$$

Definition 27. *Poission Distobution : $X \sim Poiss(\lambda)$ if we may say that*

$$p_X(x) = \frac{\lambda^k}{k!}e^{-\lambda}$$

Example 8. *LEt X be a drv as a indicator function of a set $A \in \mathcal{F}$. We can say here $E[X] = \mathbb{P}(A)$*

Example 9. $X \sim Geom(P)$

$$
\begin{aligned}
E[x] &= \sum_{n=1}^{\infty} n(1-p)^{n-1} p \\
&= p \sum_{n=1}^{\infty} n(1-p)^{n-1} \\
&= p \sum_{n=1}^{\infty} \frac{d}{d(1-p)} (1-p)^n \\
&= p \frac{d}{d(1-p)} \sum_{n=1}^{\infty} (1-p)^n \\
&= p \frac{d}{d(1-p)} \frac{1}{1-(1-p)} \\
&= p \frac{-1}{(1-(1-p)^2)} \\
&= p \frac{1}{p^2} \\
&= \frac{1}{p}
\end{aligned}
$$

12 Lecture 15

12.1 Expectation of functions

One interpretation of the expectation is as the centre of mass of the mean of a distrobution. This allows the expectation to act as a mean.

If we have a 2 dimensional random variable $Z = (X, Y)$ We can say that $E[Z] = (E[X], E[Y])$ Sometimes we tranform a ditrobution by a function $g : \mathbb{R} \to \mathbb{R}$ and we still want to find these expectations. We have been able to do this for linear functions but we may want to do this for x^2 as this is useful for variance.

Example 10. *Consider $X \sim Geom(p)$*

$$\begin{aligned}
E[X^2] &= \sum_{n=1}^{\infty} n^2(1-p)^{n-1}p \\
&= p\sum_{n=0}^{\infty} nx^{n-1} + p\sum_{n=0}^{\infty} n(n-1)x^{n-1}, x = 1-p \\
&= p\sum_{n=0}^{\infty} \frac{d}{dx}x^n + px\sum_{n=0}^{\infty} n(n-1)x^{n-2} \\
&= p\frac{d}{dx}\left(\sum_{n=0}^{\infty} x^n\right) + px\sum_{n=0}^{\infty}\frac{d^2}{dx^2}x^n \\
&= p\frac{d}{dx}\left(\frac{1}{1-x}\right) + px\frac{d^2}{dx^2}\left(\sum_{n=0}^{\infty} x^n\right) \\
&= p\frac{1}{(1-x)^2} + px\frac{2}{(1-x)^3} \\
&= \frac{1}{p} + 2\frac{1-p}{p^2} \\
&= \frac{2-p}{p^2}
\end{aligned}$$

Example 11. *Consider $X \sim Geom(p)$*

For what values of t is e^{tx} integrable (Sum for expectation exists). As such compute $E[e^{tX}]$.

$$\begin{aligned}
E[e^{tX}] &= \sum_{n=1}^{\infty} e^{tn}p(1-p)^{n-1} \\
&= pe^t\sum_{n=1}^{\infty}(e^t(1-p))^{n-1} \\
&= pe^t\sum_{k=0}^{\infty}(e^t(1-p))^{k} \\
&= \frac{pe^t}{1-e^t(1-p)} \\
&= \frac{p}{e^{-t}+p-1}
\end{aligned}$$

Not this is only defined where $e^t(1-p) < 1$ so $t < \ln\frac{1}{1-p}$

Consider for where it isnt integrable. This means that there is no expectation. Further if a random varibale does have an expectation a function of the random variable (even x^2) may not hadve a defined expectation.

In general We may say if you have a DRV X transformed by a function $g : \mathbb{R} \to \mathbb{R}$ We may say the following

$$E[g(X)] = \sum_{x \in D_X} g(x) p_X(x)$$

12.2 Variance

Variance is used to measure the dispersion of a random variable

Definition 28. *(Square integrable r.v.) : A DRV X is square integrable if X^2 is integrable.*

Definition 29. *(Variance): Consider a Square integrable random variable X. Let $\mu := E[X]$*

$$Var(x) = E[(x - \mu)^2]$$

From here we can note that

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

Example 12. *Consider $X \sim Poisson(\lambda)$*

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

Example 13. *Consider $X \sim Geom(p)$*

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

We may note that the follwwing identity holds

$$Var(ax) = a^2 Var(x)$$

We also can see then that the units of the variance are not therefore the same as the expexctation but instead is the square of the unit as such we introduce the following.

12.3 Standard Deviation

Definition 30. *Let X be a square integrable DRV then we may say the following*

$$\sigma(X) = \sqrt{Var(X)}$$

13 Multivariate Discrete Dictrobutions

Sometimes you wish to investigate a combination of thwo DRVs. The combination of the to can mbe made into one in of itself

13.1 Probability Mass functions

Definition 31. *Given two DRVs we can define the joint probability mass function as follows*

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

Or more in the proper times

$$\mathbb{P}(\{\omega \in \Omega | X(\omega) = x, Y(\omega) = y\})$$

Sometimes you wish to recover p_X from a jpoint probability distrbution. This is called the marginal probability and can be computes as follows

$$p_X(x) = \sum_{y \in D_Y} p_{X,Y}(x,y)$$

13.2 Expectation

Proposition 17. *Given two DRVs X, Y and a function $g : \mathbb{R}^2 \to \mathbb{R}$ we may say that*

$$E(g(X,Y)) = \sum_{x \in D_X} \sum_{y \in D_Y} g(x,y) p_{X,Y}(x,y)$$

This can be formally proven by constructing a new DRV $Z = g(X,Y)$. It is important to noe when constructing a probability measure on Z we are not guaranteed that g is an injection so there can be multiple pairs X, Y that can give the same value of Z

13.3 Indepedance

Definition 32. *We can say two DRVs X, Y are independant if*

$$p_{X,Y}(x,y) = p_X(x) p_Y(y)$$

The defintions of pairwise and mutual independance fall out similarly to before. It follows that if two DRVs are independant we can calculate the expectation of the prodict as follows

$$E(XY) = E(X)E(Y)$$

This iextends to the product of more than one DRV. Similarly we can get a result for variance.

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{x=1}^{n} Var(X_i)$$

Both of these have proofs which are omitted.

14 Law of Averages

Theorem 4. *Let (X_i) be a collection of pariwise independant discrete random variables all bearing the same mean and variance μ, σ. We may say then that*

$$\mathbb{P}\left(\mu - a \leq \frac{X_1 + \cdots + X_n}{n} \leq \mu + a\right) \geq 1 - \frac{\sigma^2}{a^2 n}$$

This essentially encodes the idea that with more and more trials the outcome becomes more and more clustered around the true mean the mean of outcomes becomes.

15 Covariance

Sometimes we want to investigate the how two random variable vary together. though expansion by the definition we can note

$$Var(X+Y) = Var(X) + Var(Y) + 2E[(X - E[X])(Y - E[Y])]$$

From here we can gain the following definition

Definition 33. *We can define the covariance of two DRV's as follows*

$$Cov(X,Y) = E[(X - E[X])(Y - E[Y])]$$

If we have that $Cov(X,Y) = 0$ then we say the two varibles are uncorrelated. All independant variables are uncorrelated Covariance has a few simple identities

$$Cov(X,Y) = Cov(Y,X)$$
$$Cov(X,X) = Var(X)$$

Covariance comes with a number of useful algebraic properties.

Proposition 18. *Suppose X, Y, Z are square integrable*

$$Cov(aX + bY, Z) = aCov(X,Z) + bCov(Y,Z)$$

This propertiy naturally extends as

Proposition 19.

$$Cov\left(\sum_{j=0}^{n} a_j X_j, \sum_{k=0}^{m} b_k Y_k\right) = \sum_{j=0}^{n}\sum_{k=0}^{m} a_j b_k Cov(X_j, Y_k)$$

We can also generlaise the definition of covariance from above aswell

Proposition 20.

$$Var\left(\sum_{k=1}^{n} X_k\right) = \sum_{x=k}^{n} Var(X_k) + 2\sum_{1\leq j\leq k\leq n} Cov(X_j, X_k)$$

16 Chebyshevs Inequality

16.1 Markovs Inequality

Theorem 5. *(Markovs Inequality). Let X be a non negative integrable DRV we may say*

$$\mathbb{P}(X \geq x) \leq \frac{E[X]}{x}$$

For all $x > 0$

Proof. Fix $x > 0$ then we can define a radnom variable Y as follows

$$Y = \begin{cases} x & X \geq x \\ 0 & \text{Otherwise} \end{cases}$$

It is clear that $X \geq Y$ so we may also then say $E[X] \geq E[Y]$ We can also note that the probability mass function is given as

$$\begin{aligned} p_Y(x) &= \mathbb{P}(X \geq x) \\ p_Y(0) &= \mathbb{P}(X < x) \end{aligned}$$

as such we can do the following

$$\begin{aligned} E[X] &\geq E[Y] \\ &\geq 0p_Y(0) + xp_Y(x) \\ &\geq xP(X \geq x) \end{aligned}$$

Which gives the desired inequality □

16.2 Chebyshevs Inequality