1 Probabilities

1.1 Experiments with randomness

We will use the term *experiment* in a very general way to refer to some process that produces a random outcome.
Examples: (Ask class for some first)

Here are some *discrete* examples:

- roll a die
- flip a coin
- flip a coin until we get heads

And here are some *continuous* examples:

- height of a U of A student
- random number in $[0, 1]$
- watch a radioactive substance until it undergoes a decay

These examples share the following common features: There is a procedure or natural phenomena called the *experiment*. It has a set of possible *outcomes*. There is a way to assign probabilities to sets of possible outcomes. We will call this a *probability measure*.

1.2 Outcomes and events

Definition 1. *An* **experiment** *is a well defined prodecedure or sequence of procedures that produces an* **outcome**. *The set of possible outcomes is called the* **sample space**. *We will often denote outcomes by ω and the sample space by Ω.*

Definition 2. *An* **event** *is a subset of the sample space.*

This definition will be changed when we come to the definition of a σ-field.

The next thing to define is a probability measure. Before we can do this properly we need some more structure, so for now we just make an informal definition. A probability measure is a function on the collection of events that assign a number between 0 and 1 to each event and satisfies certain properties.

NB: A probability measure is not a function on Ω.

Set notation: $A \subset B$, A is a subset of B, means ... The union $A \cup B$ of A and B is ... The intersection $A \cap B$ of A and B is ...

$\cup_{j=1}^n A_j$ is ...

$\cap_{j=1}^n A_j$ is ...

$\cap_{j=1}^\infty A_j, \quad \cup_{j=1}^\infty A_j$ are ...

Two sets A and B are **disjoint** if $A \cap B = \emptyset$. $\emptyset$ is ...

Complements: The **complement** of an event A, denoted A^c, is the set of outcomes (in Ω) which are not in A. Note that the book writes it as $\Omega \setminus A$.

De Morgan's laws:

$$\begin{aligned}(A \cup B)^c &= A^c \cap B^c \\ (A \cap B)^c &= A^c \cup B^c \\ (\cup_j A_j)^c &= \cap_j A_j^c \\ (\cap_j A_j)^c &= \cup_j A_j^c\end{aligned} \tag{1}$$

Proving set identities To prove $A = B$ you must prove two things $A \subset B$ and $B \subset A$. It is often useful to draw a picture (Venn diagram).

Example Simplify $(E \cap F) \cup (E \cap F^c)$.

1.3 Probability measures

Before we get into the mathematical definition of a probability measure we consider a bunch of examples. For an event A, $\mathbf{P}(A)$ will denote the probability of A. Remember that A is typically not just a single outcome of the experiment but rather some collection of outcomes.

What properties should P have?

$$0 \leq \mathbf{P}(A) \leq 1 \tag{2}$$

$$\mathbf{P}(\emptyset) = 0, \quad \mathbf{P}(\Omega) = 1 \tag{3}$$

If A and B are disjoint then

$$\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) \tag{4}$$

Example (uniform discrete): Suppose Ω is a finite set and the outcomes in Ω are (for some reason) equally likely. For an event A let $|A|$ denote the cardinality of A, i.e., the number of outcomes in A. Then the probability measure is given by

$$\mathbf{P}(A) = \frac{|A|}{|\Omega|} \tag{5}$$

Example (roll a die): Then $\Omega = \{1, 2, 3, 4, 5, 6\}$. If $A = \{2, 4, 6\}$ is the event that the roll is even, then $\mathbf{P}(A) = 1/2$.

Example (uniform continuous): Pick random number between 0 and 1 with all numbers equally likely. (Computer languages typically have a function that does this, for example drand48() in C. Strictly speakly they are not truly random.) The sample space is $\Omega = [0, 1]$. For an interval I contained in $[0, 1]$, it probability is its length.

More generally, the uniform probability measure on $[a, b]$ is defined by

$$\mathbf{P}([c, d]) = \frac{d - c}{b - a} \tag{6}$$

for intervals $[c, d]$ contained in $[a, b]$.

Example (uniform two dimensional): A circular dartboard is $10in$ in radius. You throw a dart at it, but you are not very good and so are equally likely to hit any point on the board. (Throws that miss the board completely are done over.) For a subset E of the board,

$$\mathbf{P}(E) = \frac{area(E)}{\pi 10^2} \tag{7}$$

Mantra: For uniform continuous probability measures, the probability acts like length, area, or volume.

Example: Roll two four-sided dice. What is the probability we get a 4 and a 3?

Wrong solution

Correct solution

Example (geometric): Roll a die (usual six-sided) until we get a 1. Look at the number of rolls it took (including the final roll which was 1). The sample space is infinite: $\Omega = \{1, 2, 3, \cdots\}$. What is the probability of it takes n rolls? This means you get $n - 1$ rolls that are not a 1 and then a roll that is a 1. So

$$\mathbf{P}(n) = \left(\frac{5}{6}\right)^{n-1} \frac{1}{6} \tag{8}$$

It should be true that if we sum this over all possible n we get 1. To check this we need to recall the geometric series formula:

$$\sum_{n=0}^{\infty} r^n = \frac{1}{1-r} \tag{9}$$

if $|r| < 1$.

Check normalization. **GAP !!!!!!!**

Compute $\mathbf{P}$(number of rolls is odd). **GAP !!!!!!!**

End of August 23 lecture

We now return to the definition of a probability measure. We certainly want it to be additive is the sense that if two events A and B are disjoint, then

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

Using induction this property implies that if $A_1, A_2, \cdots A_n$ are disjoint ($A_i \cap A_j = \emptyset, i \neq j$), then

$$\mathbf{P}(\cup_{j=1}^n A_j) = \sum_{j=1}^n \mathbf{P}(A_j)$$

It turns out we need more than this to get a good theory. We need a infinite version of the above. If we require that the above holds for *any* infinite disjoint union, that is too much and the result is there are reasonable probability measures that get left out. The correct definition is that this property holds for *countable* unions.

GAP !!!!!!!!!!!!!!!!!!!!! Review what countable means.

The property we will require for probability measures is the following. Let A_n be a sequence of disjoint events, i.e., $A_i \cap A_j = \emptyset$ for $i \neq j$. Then we require

$$\mathbf{P}(\cup_{j=1}^\infty A_j) = \sum_{j=1}^\infty \mathbf{P}(A_j)$$

Until now we have defined an event to just be a subset of the sample space. If we allow all subsets of the sample space to be events, we get in trouble. Consider the uniform probability measure $\mathbf{P}$ on $[0,1]$ So for $0 \le a < b \le 1$, $\mathbf{P}([a,b]) = b - a$. We would like to extend the definition of $\mathbf{P}$ to all subsets of $[0,1]$ in such a way that (10) holds. Unfortunately, one can prove that this cannot be done. The way out of this mess is to only define $\mathbf{P}$ on a *subcollection* of the subsets of $[0,1]$. The subcollection will contain all "reasonable" subsets of $[0,1]$, and it is possible to extend the definition of $\mathbf{P}$ to the subcollection in such a way that (10) holds provided all the A_n are in the subcollection. The subcollection needs to have some properties. This gets us into the rather technical definition of a σ-field. The concept of a σ-field plays an important role in more advanced probability. It will not play a major role in this course. In fact, after we make this definition we will not see σ-fields again until near the end of the course.

Definition 3. *Let Ω be a sample space. A collection $\mathcal{F}$ of subsets of Ω is a σ-field if*

1. $\Omega \in \mathcal{F}$
2. $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$
3. $A_n \in \mathcal{F}$ *for* $n = 1, 2, 3, \cdots \quad \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$

The book calls a σ-field the "event space." I will not use this terminology. From now on, when I use the term event I will mean not just a subset of the sample space but a subset that is in $\mathcal{F}$.

Roughly speaking, a σ-field has the property that if you take a countable number of events and combine them using a finite number of unions, intersections and complements, then the resulting set will also be in the σ-field. As an example of this principle we have

Theorem 1. *If $\mathcal{F}$ is a σ-field and $A_n \in \mathcal{F}$ for $n = 1, 2, 3, \cdots$, then $\cap_{n=1}^{\infty} A_n \in \mathcal{F}$.*

Proof. **GAP !!!!!!!!!!!!** $\square$

We can finally give the mathematical definition of a probability measure.

Definition 4. *Let $\mathcal{F}$ be a σ-field of events in Ω. A probability measure on $\mathcal{F}$ is a real-valued function $\mathbf{P}$ on $\mathcal{F}$ with the following properties.*

1. $\mathbf{P}(A) \geq 0$, *for* $A \in \mathcal{F}$.

2. $\mathbf{P}(\Omega) = 1$, $\mathbf{P}(\emptyset) = 0$.

3. *If* $A_n \in \mathcal{F}$ *is a disjoint sequence of events, i.e.,* $A_i \cap A_j = \emptyset$ *for* $i \neq j$, *then*

$$\mathbf{P}(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mathbf{P}(A_n) \tag{10}$$

1.4 Properties of probability measures

We start with a bit of terminology. If Ω is a set (the probability space), $\mathcal{F}$ is a σ-algebra of subsets in Ω (the events), and $\mathbf{P}$ is a probability measure on $\mathcal{F}$, then we refer to the triple $(\Omega, \mathcal{F}, \mathbf{P})$ as a *probability space*.

The next theorem gives various formulae for computing with probability measures.

Theorem 2. *Let* $(\Omega, \mathcal{F}, \mathbf{P})$ *be a probability space.*

1. $\mathbf{P}(A^c) = 1 - \mathbf{P}(A)$ *for* $A \in \mathcal{F}$.

2. $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$ *for* $A, B \in \mathcal{F}$.

3. $\mathbf{P}(A \setminus B) = \mathbf{P}(A) - \mathbf{P}(A \cap B)$. *for* $A, B \in \mathcal{F}$.

4. *If* $A \subset B$, *then* $\mathbf{P}(A) \leq \mathbf{P}(B)$. *for* $A, B \in \mathcal{F}$.

5. *If* $A_1, A_2, \cdots, A_n \in \mathcal{F}$ *are disjoint, then*

$$\mathbf{P}(\bigcup_{j=1}^{n} A_j) = \sum_{j=1}^{n} \mathbf{P}(A_j) \tag{11}$$

Proof. Prove some of them. **GAP !!!!!!!!!!!!!!!!!!!!!** □

We have given a mathematical definition of probability measures and proved some properties about them. We now try to give some intuition about what the probability measure tells you. To be concrete we consider an example. We roll two four-sided dice and look at the probabilty their sum is less than or equal to 3. It is 3/16. Now suppose we repeat the experiment a large number of times, say $1,000$. Then we expect that out of these $1,000$ there will be approximately $3000/16$ times that we get a sum less than or equal to 3. Of course we don't expect to get exactly this many occurences. But if we do the experiment N times and look at the number of times we get a sum less than or equal to 3 divided by N, the number of times we did the experiment, then this ratio should converge to $3/16$ as N goes to infinity. More generally, if A is some event and we do the experiment N times, the we should have

$$\lim_{N\to\infty} \frac{number\ of\ outcomes\ in\ A}{N} = \mathbf{P}(A)$$

This statement is not a mathematical theorem at this point, but we will eventually make it into one. This view of probability is often called the "frequentist view."

1.5 Discrete probability measures

If Ω is countable, we will call a probability measure on Ω discrete. In this setting we can take $\mathcal{F}$ to just be all subsets of Ω. If E is an event, then it is countable and so we can write it as the countable disjoint union of sets with just one element in each of them. Then by the definition of a probability measure, $\mathbf{P}(E)$ is given by summing up the probabilities of each of the outsome in E. So in this setting $\mathbf{P}$ is completely determined by the numbers $\mathbf{P}(\{\omega\})$ were ω is an outcome in Ω. We emphasize that this is not true for uncountable Ω.

Proposition 1. *Let Ω be countable. Write is as $\Omega = \{\omega_1, \omega_2, \omega_3, \cdots\}$. Let p_n be a sequence of non-negative numbers such that*

$$\sum_{n=1}^{\infty} p_n = 1$$

Define $\mathcal{F}$ to be all subsets of Ω. For an event $E \in \mathcal{F}$, define its probability by

$$\mathbf{P}(E) = \sum_{n:\omega_n \in E} p_n$$

Then $\mathbf{P}$ is a probability measure on $\mathcal{F}$.

Proof. The first two properties of a probability measure are obvious. The countable additivity property comes down to the theorem that says you can rearrange an absolutely convergent series and not change its value. $\square$

We have already seen examples of such probability measures. For following give Ω and the p_n.

- Roll a die.
- Roll a die until we get a 6.
- Flip a coin n times and look at total number of heads.

1.6 Conditional probability

Conditional probability is a very important idea in probability. We start with the formal mathematical definition and will then explain why this is a good definition.

Definition 5. *If A and B are events and $\mathbf{P}(B) > 0$, then the (conditional) probability of A given B is denoted $\mathbf{P}(A|B)$ and is given by*

$$\mathbf{P}(A|B) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)} = \frac{intersection}{given} \tag{12}$$

Example: Pick a real random number X uniformly from $[0, 10]$. If $X > 6.5$, what is the probability $X > 7.5$.

Theorem 3. *Let $\mathbf{P}$ be a probability measure, B an event ($B \in \mathcal{F}$) with $\mathbf{P}(B) > 0$. For events A ($A \in \mathcal{F}$)), define*

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

Then Q is a probability measure on $(\Omega, \mathcal{F})$.

Proof. Prove countable additivity. **GAP !!!!!!!!!!!!!!** $\square$

To motivate the definition consider the following example. Suppose we roll two four-sided dice, but we only “keep” the roll if the sum is less than or equal to 3. For this “conditional” experiement we ask what is the probability that we do not get a 2 on either die. We can look at this from a frequentist viewpoint. Suppose we roll the two dice a large number of times, say N. The probability of the sum being less than or equal to 3 is 3/16, so we expect to get approximately $3N/16$ that have a sum less than or equal to 3. How many of them do not have a 2? If the sum is less than or equal to 2 and they do not have have a 2, then the roll is two 1's. Note that we are taking the intersection here. The probability of this intersection event is 1/16. So the number of rolls that we "count" which also have no 2's should be approximately $N/16$. So the fraction of the rolls that count which has no 2 is approximately

$$\frac{\frac{3N}{16}}{\frac{N}{16}} = \frac{\frac{3}{16}}{\overline{16}} = \frac{\mathbf{P}(intersection)}{\mathbf{P}(given)}$$

TREES ¡¡¡¡¡¡¡¡¡¡¡¡¡¡¡,,