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 $[0, 1]$
- the time it takes until a radioactive substance 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 procedure or sequence of procedures that produces an* **outcome**. *The set of possible outcomes is called the* **sample space**. *We will typically denote an individual outcome 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 that every element of A is also in B. The union $A \cup B$ of A and B is the of all elements that are in A or B, including those that are in both. The intersection $A \cap B$ of A and B is the set of all elements that are in both of A and B.

$\cup_{j=1}^n A_j$ is the set of elements that are in at least one of the A_j.

$\cap_{j=1}^n A_j$ is the set of elements that are in all of the A_j.

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

Two sets A and B are **disjoint** if $A \cap B = \emptyset$. $\emptyset$ denotes the empty set, the set with no elements.

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:

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

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 (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 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 (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 speaking 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). There are two possible choices for the sample space. We could consider an outcome to be a sequence of rolls that end with a 1. But if all we care about is how many rolls it takes, then we can just take the sample space to be $\Omega = \{1, 2, 3, \cdots\}$, where the outcome corresponds to how many rolls it takes. 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 !!!!!!!**

Geometric: We generalize the above example. We start with a simple experiment with two outcomes which we refer to as *success* and *failure*. For the simple experiment the probability of success is p, and the probability of failure is $1-p$. The experiment we consider is to repeat the simple experiment until we get a success. Then we let n be the number of times we repeated the simple experiment. The sample space is $\Omega = \{1, 2, 3, \cdots\}$. The probability of a single outcome is given by

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

You can use the geometric series formula to check that if you sum this probability over $n = 1$ to ∞, you get 1.

End of August 23 lecture

We now return to the definition of a probability measure. We certainly want it to be additive in 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.

We briefly review what countable means. A set is countable is there is a bijection (one-to-one and onto map) between the set and the natural numbers. So any sequence is countable and every countable set can be written as a sequence.

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 $0 \leq a < b \leq 1$, $\mathbf{P}([a, b]) = b - a$. 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 !!!!!!!!!!!!** □

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*

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

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 probability 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 occurrences. 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 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 $X > 7.5$.

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" experiment 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

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

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 !!!!!!!!!!!!!!** □

Trees

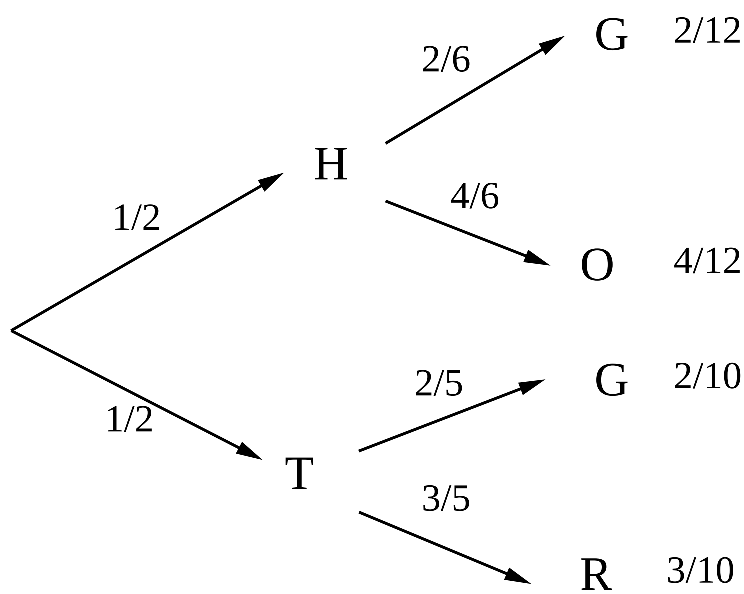

Figure 1: Tree for the example with flipping a coin and then removing some balls from the urn.

Trees are a nice tool in solving problems involving conditional probability. We can rewrite the definition of conditional probability as

$$\mathbf{P}(A \cap B) = \mathbf{P}(A|B)\mathbf{P}(B) \tag{13}$$

In some experiments the nature of the experiment means that we know certain conditional probabilities. If we know $\mathbf{P}(A|B)$ and $\mathbf{P}(B)$, then we can use the above to compute $\mathbf{P}(A \cap B)$. We illustrate this with an example.

Example: An urn contains 3 red balls, 2 green balls and 4 orange balls. The balls are identical except for their color. We flip a fair coin. If it is heads, we remove all the red balls. If it is tails, we remove all the orange balls. Then we draw one ball from the urn and look at its color. We take an outcome to be both the result of the coin flip and the color of the ball. So

$$\Omega = \{(H,G),(H,O),(H,R),(T,G),(T,O),(T,R)\} \tag{14}$$

Figure 1 shows the tree to analyze this example. Consider following the tree from left to right along the topmost branches. This corresponds to a coin flip of heads, followed by drawing a green ball. The probability of heads is 1/2. Given that the flip was heads the probability of green is 2/6. Multiplying these two numbers gives the probability of the outcome (H, G), i.e, 2/12.

End of August 25 lecture

1.6 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 countable additivity property of a probability measure, $\mathbf{P}(E)$ is given by summing up the probabilities of each of the outcomes 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 it 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. □

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.

1.7 Independence

In general $\mathbf{P}(A|B) \neq \mathbf{P}(A)$. Knowing that B happens changes the probability that A happens. But sometimes it does not. Using the definition of $\mathbf{P}(A|B)$, if $\mathbf{P}(A) = \mathbf{P}(A|B)$ then

$$\mathbf{P}(A) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)} \tag{15}$$

i.e., $\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B)$. This motivate the following definition.

Definition 6. *Two events are independent if*

$$\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B) \tag{16}$$

Caution: Independent and disjoint are two quite different concepts. If A and B are disjoint, then $A \cap B = \emptyset$ and so $\mathbf{P}(A \cap B) = 0$. So they are also independent only if $\mathbf{P}(A) = 0$ or $\mathbf{P}(B) = 0$.

Example Roll 2 four-sided dice. Define four events:

$$\begin{aligned} A &= \textit{first roll is a } 2 \\ B &= \textit{second roll is a } 3 \\ C &= \textit{sum is } 6 \textit{ or } 8 \\ D &= \textit{product is odd} \end{aligned} \tag{17}$$

Show that C and D are independent and most of the other pairs are not. **GAP !!!!!!!!!!!!!!!!**

Using and abusing the definition: Often we use the idea of independent to figure out what the probability measure is. We have already done this implicitly in some examples. In the "roll a die until we get a 6" example, we compute things like the probability that it takes 3 rolls by assuming the roles were independent: $\frac{5}{6} \times \frac{5}{6} \times \frac{1}{6}$. But if we are given P, then determining if two events are independent is just a matter of computation, not a philosophical question. In the previous example with two dice one might think that C and D should influence each other and so are not independent. But the computation shows they are.

Theorem 4. *If A and B are independent events, then*

1. *A and B^c are independent*
2. *A^c and B are independent*
3. *A^c and B^c are independent*

In general A and A^c are not independent.

The notion of independence can be extended to more than two events.

Definition 7. *Let $A_1, A_2, \cdots, A_n$ be events. They are independent if for all subsets I of $\{1, 2, \cdots, n\}$ we have*

$$\mathbf{P}(\cap_{i \in I} A_i) = \prod_{i \in I} \mathbf{P}(A_i) \tag{18}$$

They are just pairwise independent if $\mathbf{P}(A_i \cap A_j) = \mathbf{P}(A_i)\mathbf{P}(A_j)$ for $1 \leq i < j \leq n$.

Obviously, independent implies pairwise independent. But it is possible to have events that are pairwise independent but not independent. One of the homework problems will give an example of this.

1.8 The partition theorem and Bayes theorem

We start with a definition.

Definition 8. *A partition is a finite or countable collection of events B_j such that $\Omega = \cup_j B_j$ and the B_j are disjoint, i.e., $B_i \cap B_j = \emptyset$ for $i \neq j$.*

Theorem 5. *(Partition theorem) Let $\{B_j\}$ be a partition of Ω. Then for any event A,*

$$\mathbf{P}(A) = \sum_j \mathbf{P}(A|B_j)\, \mathbf{P}(B_i) \tag{19}$$

Proof. We start with the set identity

$$A = A \cap (\cup_j B_j) = \cup_j (A \cap B_j) \tag{20}$$

So

$$\mathbf{P}(A) = \mathbf{P}(\cup_j (A \cap B_j)) = \sum_j \mathbf{P}(A \cap B_j) \tag{21}$$

where the last inequality holds since the events $A \cap B_j$ are disjoint. We can write $\mathbf{P}(A \cap B_j)$ as $\mathbf{P}(A|B_j)\mathbf{P}(B_j)$, so we obtain the equation in the theorem. □

The formula in the theorem should be intuitive.

Example A box has 4 red balls. We flip a fair coin until we get a 6. Let N be the number of flips it takes (so $N \geq 1$). We add N green balls to the box. Then we draw a ball from the box. Find the probability the ball drawn is red.

GAP !!!!!!!!!!!!!!! Solution

Example We return to the example above with a box containing 4 red balls to which we add N green balls where N is the number of flips of a fair coin to get a 6.
(a) Given that $N = 3$, find the probability the ball drawn is red.
(b) If the ball drawn is red, what is $P(N = 3)$?

The first question is trivial. The second one is often called a Bayes theorem problem. In this problem it is trivial to find the probability of the ball being a certain color if we know the value of N. But if we interchange these events and ask for the probability that N is a certain value given the color of the ball drawn, that is a more complicated question. We have to use the definition of conditional probability and the partition theorem. You can write down a complicated theorem/formula (Bayes theorem) that can be used to do (b), but we won't. All you really need is the definition of conditional probability and the partition theorem.

GAP !!!!!!!!!!!!!!! Solution

Often the hardest part of using the partition theorem is figuring out what partition to use. The following example illustrates this point.

Example: We flip a fair coin n times. When we get k heads in a row, this is called a *run* of k heads. We want to find the probability that we do not get a run of 3 heads in the n flips. Let u_n be this probability. We will find a recursion relation for u_n as a function of n. Note that $u_1 = 1$ and $u_2 = 1$. We will look at the first three flips to define our partition. Our partition has 4 events. We denote them by just T,HT,HHT, HHH. T is the event that the first flip is tails. HT is the event that the first flip is heads and the second is tails. HHT is the event that the first two flips are heads and the third flip is T. HHH is ... Let E be the event that there is no run of 3 heads in the n flips. The partition theorem says

$$\begin{aligned}\mathbf{P}_n(E) &= \mathbf{P}_n(E|T)\,\mathbf{P}_n(T) + \mathbf{P}_n(E|HT)\,\mathbf{P}_n(HT) \\ &+ \mathbf{P}_n(E|HHT)\,\mathbf{P}_n(HHT) + \mathbf{P}_n(E|HHH)\,\mathbf{P}_n(HHH)\end{aligned}$$

We have added a subscript n to $\mathbf{P}$ to indicate that this is the probability measure for the experiment of flipping the coin n times. Some of these probabilities are easily evaluated : $\mathbf{P}_n(T) = 1/2, \mathbf{P}_n(HT) = 1/4, \cdots$, Now consider $\mathbf{P}_n(E|T)$. Since the first flip is T, the first flip will not be part of a run of 3 heads. So $\mathbf{P}_n(E|T) = \mathbf{P}_{n-1}(E)$. Similarly, $\mathbf{P}_n(E|HT) = \mathbf{P}_{n-2}(E)$ and $\mathbf{P}_n(E|HHT) = \mathbf{P}_{n-3}(E)$. Note that $\mathbf{P}_n(E|HHH) = 1$. Letting p_n denote $\mathbf{P}_n(E)$ we get a recursion relation:

$$p_n = \frac{1}{2}p_{n-1} + \frac{1}{4}p_{n-2} + \frac{1}{8}p_{n-3} + \frac{1}{8}$$

The first few p_n are easily seen to be $p_1 = p_2 = 0$, $p_3 = 1/8$.

Example - the Monty Hall problem There are three doors. Behind one is a new car. Behind the other two are goats. You get to pick a door. The host (Monty Hall) knows where the car is. After you have picked, he opens a door that has a goat behind it. You are give the option of sticking with your original choice or switching to the other unopened door. What is you best strategy? Here are three strategies:

1. Stick with your original choice.
2. Always switch
3. Flip a coin. If it is heads then switch. It tails, stick.

Compute the probability of winning a car for each strategy.

End of August 30 lecture

GAP !!!!!!!!!!!!!!! Solution

1.9 Continuity of P

A sequence of events A_n is said to be *increasing* if $A_1 \subset A_2 \subset A_3 \cdots A_n \subset A_{n+1} \cdots$. In this case we can think of $\cup_{n=1}^{\infty} A_n$ as the "limit" of A_n.

Similarly, A_n is said to be *decreasing* if $A_1 \supset A_2 \supset A_3 \cdots A_n \supset A_{n+1} \cdots$. In this case we can think of $\cap_{n=1}^{\infty} A_n$ as the "limit" of A_n.

The following theorem says that probability measures are continuous in some sense.

Theorem 6. *Let* $(\Omega, \mathcal{F}, \mathbf{P})$ *be a probability space. Let* A_n *be an increasing sequence of events. Then*

$$\mathbf{P}(\cup_{n=1}^{\infty} A_n) = \lim_{n\to\infty} \mathbf{P}(A_n) \tag{22}$$

If A_n *is a decreasing sequence of events, then*

$$\mathbf{P}(\cap_{n=1}^{\infty} A_n) = \lim_{n\to\infty} \mathbf{P}(A_n) \tag{23}$$

Example Roll a die forever. What is Ω ? Let A_n be the event that there is no 1 in the first n rolls. Then A_n is a decreasing sequence of events. We have $\mathbf{P}(A_n) = \left(\frac{5}{6}\right)^n$. So $\lim_{n\to\infty} \mathbf{P}(A_n) = 0$. So the theorem says

$$\mathbf{P}(\cap_{n=1}^{\infty} A_n) = 0 \tag{24}$$

The event $\cap_{n=1}^{\infty} A_n$ is the event that for every n there is no 1 in the first n rolls. So it is the event that we never get a 1. The probability this happens is 0, i.e., the probability we eventually get a 1 is 1.