

Discrete Probability

Discrete Probability

**the sum of $U = 0$

so out of a 2 bit universe... everything has to equal 1

00 01 10 11 the probability

$1/2$ = the probability of 00

$1/8$ the probability of 01

$1/4$ the probability of 10

$1/8$ the probability of 11

$$1/8 + 1/2 + 1/4 + 1/8 = 1$$

looking at example we set $1/u$

↗ U: finite set (e.g. $U = \{0,1\}^n$)

$$\{0,1\}^2 = \{00, 01, 10, 11\}$$

Def: Probability distribution P over U is a function $P: \underline{U} \rightarrow \underline{[0,1]}$

such that $\sum_{x \in U} P(x) = 1$

U: finite set (e.g. $U = \{0,1\}^n$)

$$\{0,1\}^2 = \{00, 01, 10, 11\}$$

Def: Probability distribution P over U is a function $P: U \rightarrow [0,1]$

such that $\sum_{x \in U} P(x) = 1$

$$P: \begin{array}{cccc} 00 & 01 & 10 & 11 \\ \downarrow & \downarrow & \downarrow & \downarrow \end{array}$$

Examples:

1. Uniform distribution: for all $x \in U$: $P(x) = 1/|U|$

U: finite set (e.g. $U = \{0,1\}^n$)

$$\{0,1\}^2 = \{00, 01, 10, 11\}$$

Def: Probability distribution P over U is a function $P: U \rightarrow [0,1]$

such that $\sum_{x \in U} P(x) = 1$

$$P: \begin{array}{cccc} 00 & 01 & 10 & 11 \\ \downarrow & \downarrow & \downarrow & \downarrow \\ 1/4 & 1/4 & 1/4 & 1/4 \end{array}$$

Examples:

1. Uniform distribution: for all $x \in U$: $P(x) = 1/|U|$

2. Point distribution at x_0 : $P(x_0) = 1$, $\forall x \neq x_0$: $P(x) = 0$

1. Uniform distribution: for all $x \in U$: $P(x) = 1/|U|$

2. Point distribution at x_0 : $P(x_0) = 1$, $\forall x \neq x_0$: $P(x) = 0$

For all

Distribution vector: $(P(000), P(001), P(010), \dots, P(111))$

Events


- For a set $A \subseteq U$: $\Pr[A] = \sum_{x \in A} P(x) \in [0,1]$

note: $\Pr[U]=1$

- The set A is called an event

Example: $U = \{0,1\}^8$

- $A = \{ \text{all } x \text{ in } U \text{ such that } \text{lsb}_2(x)=11 \} \subseteq U$

for the uniform distribution on $\{0,1\}^8$: $\Pr[A] =$ 

Over the years many natural cryptographic constructions were found to be insecure. In response, modern cryptography was developed as a rigorous science where constructions are always accompanied by a proof of security. The language used to describe security relies on discrete probability. In this segment and the next, I'll give a brief overview of discrete probability, and I point to this Wiki books article over here for a longer introduction. Discrete probability is always defined over a universe which I'll denote by U and this universe in our case is always going to be a finite set. In fact very commonly our universe is going to be simply the set of all n bit strings which here is denoted by $\{0,1\}^n$. So for example the set $\{0,1\}^2$ is the set of all two bit strings which happens to be the string zero, zero, zero one, One, zero, and one, one. So there are four elements in this set, and more generally in the set $\{0,1\}^N$, there are 2^N elements. Now a probability distribution over this universe U is simply a function which I'll denote by P , and this function, what it does, is it assigns to every element in the universe a number between zero and one. And this number is what I'll call the weight or the probability of that particular element in the universe

***Now there's only one requirement on this**

function P , and that is that the sum of all the weights, sum up to one. That is, if I sum the probability of all elements x in the universe, what I end up with is the number one. So let's look at a very simple example looking back to our 2-bit universe. So 0001, ten and eleven And you can consider the following probability distribution Which, for example, assigns to the element 00, the probability one half. The elements 01, we assign the probability $1/8$ th, to ten we assign the probability one quarter and to eleven we assign the probability $1/8$ th. Okay we can see that if we sum up these numbers in fact we get one which means that this probability P is in fact the probability distributio N . Now what these number mean

is that if I sample from this probability distribution I'll get the string 00 with probability one half I'll get the string 01 with probability 1/8th and so on and so forth. So now that we understand what a probability distribution is, let's look at two classic examples of probability distributions. The first one is what's called the uniform distribution. **

U: finite set (e.g. $U = \{0,1\}^n$) $\{0,1\}^2 = \{00, 01, 10, 11\}$

Def: Probability distribution P over U is a function $P: U \rightarrow [0,1]$
 such that $\sum_{x \in U} P(x) = 1$

**The uniform distribution assigns to every element in the universe, exactly the same weight. I'm gonna use U between two bars to denote the size of the universe U . That is the number of elements in the universe, and since we want the sum of all the weights to sum out to one, and we want all these weights to be equal, what this means is that for every element x in the universe, we assign a probability of one over U . So in particular if we look at our example, the uniform distribution and the set of two strings, would simply assign one-quarter the weight, one-quarter to each one of these strings. And clearly that, the sum of all the weights sums up to one. Well again, what this means is that if I sample at random from this distribution, I'll get a uniform sample across all our 2-bit strings. So all of these 4-bit strings are equally likely to be sampled by this distribution. **Another distribution that's very common is what's called a point distribution at the point x_0 . And what this point distribution does is basically it puts all the weight on a single point, namely x_0 . So here we assign to the point x_0 all the weight, one. And then to all other points in the universe, we assign the weight zero. And by the way, I want to point out that this, inverted, $\forall x \neq x_0, P(x) = 0$ should be read as, for all. So all this says is, that for all x that are not equal to x_0 , the probability of that x is zero. So again going back to our example a point distribution for example, that would put

all its mass on the string 1-0, would assign probability one to the string 1-0 and zero to all other strings. So now if I sample from this distribution pretty much I'm always guaranteed to always sample the string 1-0 and never sample any of the other strings. So now we know what a distribution is, and I just want to make one last point, and that is that because this universe U is always gonna be a finite set up for us, we can actually write down the weights that the distribution assigns to every element in U , and represent the entire distribution as a vector.

U : finite set (e.g. $U = \{0,1\}^n$)

$$\{0,1\}^2 = \{00, 01, 10, 11\}$$

Def: Probability distribution P over U is a function $P: U \rightarrow [0,1]$

such that $\sum_{x \in U} P(x) = 1$

$$P: \begin{array}{c} 00 \\ \downarrow \\ 1/4 \end{array} \quad \begin{array}{c} 01 \\ \downarrow \\ 1/4 \end{array} \quad \begin{array}{c} 10 \\ \downarrow \\ 1/4 \end{array} \quad \begin{array}{c} 11 \\ \downarrow \\ 1/4 \end{array}$$

Examples:

1. Uniform distribution: for all $x \in U$: $P(x) = 1/|U|$
2. Point distribution at x_0 : $P(x_0) = 1$, $\forall x \neq x_0$: $P(x) = 0$

**So, here for example, if you look at the universe of an all 3-bit strings, we can literally write down the ways that the distribution assigns to the string 000, then the way that distribution assigns to the string 001 And so on, and so forth. We you can see that we can write this as a vector, in this case it will be a vector of dimension eight, there will be, there are eight strings of 3-bits as a result basically the entire distribution is captured by this vector of eight real numbers, in the range of all zero to one. The next thing I wanna do is define the concept of an event. So consider a subset A of our universe And I, I'll define the probability of the subsets to be simply the sum of the weights of all the elements in the set A . In other words, I'm summing over all X and A , the weights of these elements X in the set A , Now because the sum over the entire universe of all weights needs to be one. This means that if we sum, well if you look at the probability of the entire universe, basically we get one. And if we look at the probability of a subset of the universe, we're gonna get some number in the interval zero to one And we say that the probability of this set A , is the sum which is a number between zero and one. And I'll tell you that a subset A of the universe is called an event. And the probability of the set A is called the probability of that event. So let's look at a simple example.

Events


- For a set $A \subseteq U$: $\Pr[A] = \sum_{x \in A} P(x) \in [0,1]$

note: $\Pr[U]=1$

- The set A is called an event

Example: $U = \{0,1\}^8$

- $A = \{ \text{all } x \text{ in } U \text{ such that } \text{lsb}_2(x)=11 \} \subseteq U$

for the uniform distribution on $\{0,1\}^8$: $\Pr[A] =$ 

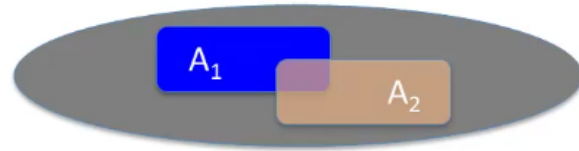
***So suppose we look at

the universe u , which consists of all 8-bit strings, right? So the size of this universe u is 256 because there are 256 8-bit strings. Essentially we're looking at all byte values, all 256 possible byte values. Now let's define the following event. Basically the event is gonna contain all bytes so all extremes in the universe such that the two least significant bits of the byte happens to be eleven. So for example, if we look at 01011010 that's an element in the universe that's not in the set A , but if we choose a zero to a one. Then that's an element of the universe which gives in our set A . And now let's look at the uniform distribution over the universe U and let me ask you what is the probability of the, of the event A ? So what is the probability that when we choose a random byte, the two least significant bits of that byte happens to be one, one? Well the answer is one-fourth, and the reason that's true is because it's not too difficult to convince yourself that of the 256 eight bit strings, exactly 64 of them, one quarter of them, end in one, one. And the probability of each string is, we're looking at the uniform distribution or probability of each string is exactly one over the size of the universe, mainly one over 256. And the product of these, you know, 64 elements, each one has weight one over 256 is exactly one-fourth, which is the probability of the event A that we're looking at. So a very simple bound on the probability of events is called the union bound

The union bound

- For events A_1 and A_2

$$\Pr[A_1 \cup A_2] \leq \Pr[A_1] + \Pr[A_2]$$



Dan Boneh

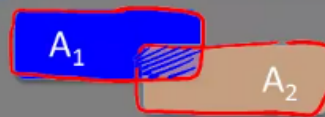
***. So imagine we have two events a_1 and a_2 . So these are both subsets of some universe U and we wanna know what is the probability that either A_1 occurs, or A_2 occurs. In other words, what is the probability of the union of these two events? This little U here denotes the union of the two sets. So the union bound tells us that the probability that either A_1 occurs or A_2 occurs is basically less than the sum of the two probabilities. And that's actually quite easy to see. So simply look at this picture here, you can see that when we look at, at the sum of the two probabilities, we're basically summing the probability of all the elements in A_1 , all the elements in A_2 . And you realized, we kind of double-summed these elements in the intersection. They get summed twice here on the right hand side. And as a result, the sum of the two probabilities is going to be larger or larger than or equal to, the actual probability of the union of A_1 and A_2 . So that's the classic union bound. And in fact I'll tell you that if the two events are disjoint, in other words they're intersection is empty, in that case if we look at the sum, at the probability that either A_1 or A_2 happens, that's exactly equal to the sum of the two probabilities. Okay?

The union bound

- For events A_1 and A_2 $\subseteq U$

$$\Pr[A_1 \cup A_2] \leq \Pr[A_1] + \Pr[A_2]$$

$$A_1 \cap A_2 = \emptyset \Rightarrow \Pr[A_1 \cup A_2] = \Pr[A_1] + \Pr[A_2]$$



***So we'll use these facts here and there throughout the course. So just to be clear, the inequality holds always. But when the two events are disjoint, then in fact we get an equality over here. So let's look at a simple example. Suppose our event A_1 is the set of all n -bit strings that happen to end in 1-1. And suppose A_2 is the set of all n -bit strings that happen to begin with 1-1. Okay, so N thinks of it as H or some large number and I'm asking that what is the probability that either a one happens or a two happens, In other words if I sample uniformly from the universe U , what is the probability that either the least significant bits are one, one or the most significant digits are one, one. Well as we said that's basically the probability of the union of A_1 and A_2 . We know that the probability of each one of these events is one-quarter by what we just did previous slide. And therefore by the union the probability of the $A_1 \cup A_2$ is, you know, a quarter of the probability of A_1 , plus the probability of A_2 , which is a quarter plus a quarter. And we just proved that the probability of seeing 1-1 in the most significant bit, or 1-1 least significant bit, is less than one-half. So

The union bound

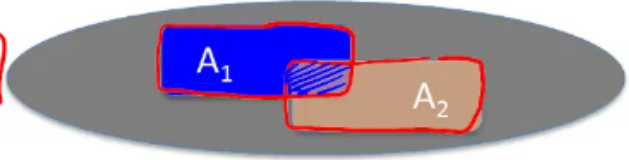
- For events A_1 and A_2 $\subseteq U$

$$\Pr[A_1 \cup A_2] \leq \Pr[A_1] + \Pr[A_2]$$

$$A_1 \cap A_2 = \emptyset \Rightarrow \Pr[A_1 \cup A_2] = \Pr[A_1] + \Pr[A_2]$$

Example:

$$A_1 = \{ \text{all } x \text{ in } \{0,1\}^n \text{ s.t. } \text{lsb}_2(x) = 11 \} \quad ; \quad A_2 = \{ \text{all } x \text{ in } \{0,1\}^n \text{ s.t. } \text{msb}_2(x) = 11 \}$$



$$\Pr[\text{lsb}_2(x) = 11 \text{ or } \text{msb}_2(x) = 11] = \Pr[A_1 \cup A_2] \leq \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

**that's a simple example of how we might go about using the Union Bound to bound the probability that one of two events might happen. The next concept we need to define, is what's called a random variable. Now, random variables are fairly intuitive objects. But unfortunately the formal definition of a random variable can be a little confusing. So what I'll do is, I'll give an example, and hopefully that will be clear enough.

Random Variables

Def: a random variable X is a function $X: U \rightarrow V$

Example: $X: \{0,1\}^n \rightarrow \{0,1\} \quad ; \quad X(y) = \text{lsb}(y) \in \{0,1\}$

So formally, a random variable denoted say, by X . Is a function, from the universe into some set V And we say that this set V is where the random variable takes its values. So let's look at a particular example. So suppose we have a random variable x And this random variable maps into the set 01 . So the values of this random variable are going to be either zero or one. So, one bit, basically. Now, this random variable maps our universe, which is the center of all end bit binary strings, 01 to the end And how does it do it?

Random Variables

Def: a random variable X is a function $X:U \rightarrow V$

Example: $X: \{0,1\}^n \rightarrow \{0,1\}$; $X(y) = \text{lsb}(y) \in \{0,1\}$

Well, given a particular sample in the universe, a particular end-bit string y . What the random variable will do is simply output the least significant bit of y . And that's it. That's the whole random variable. So, now let me ask you.

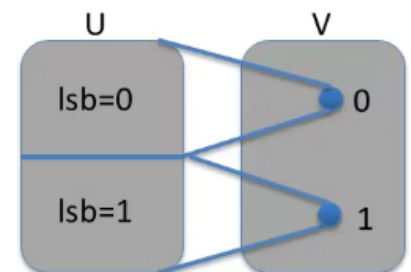
Random Variables

Def: a random variable X is a function $X:U \rightarrow V$

Example: $X: \{0,1\}^n \rightarrow \{0,1\}$; $X(y) = \text{lsb}(y)$ $\in \{0,1\}$

For the uniform distribution on U :

$$\Pr[X=0] = 1/2, \quad \Pr[X=1] = 1/2$$



Suppose we look at a uniform distribution on the set zero one to the end. Let me ask you what is the probability that this random variable output zero and what is the probability that a random variable outputs one? Well you can see that the answers are half and half. Well let's just lead them through why that's the case. So here we have a picture showing the universe and the possible alpha space. And so in this case the variable can output a zero or a one. When there is a variable output zero, there is a variable output zero when the sample in the universe happens to be, to have its least inefficient bid be set to zero. In variable one, output one When the sample in the universe happens to have its least significant bit set to one. Well, if choose strings

uniformly at random, the probability that we choose a string that has its least significant bits set to zero is exactly one half. Which the random variable output zero with a probability of exactly one-half. Similarly, if we choose a random end-bit string, the probability that the least significant bit is equal to one is also one-half. And so we say that the random variable output's one, also with exactly probability of one-half. Now, more generally, if we have a random variable taking values in a certain set v , then this random variable actually induces a distribution on this set v . And here, I just wrote a, kind of a, in symbols, what this distribution means. But it's actually very easy to explain. Essentially, what it says is that the variable outputs v . Basically, with the same probability that if we sample a random element in the universe, and, and then we apply the function x .

more generally,

rand. var. X induces a distribution on V : $\Pr[X=v] := \Pr[X^{-1}(v)]$

Dan Bo

course)

class

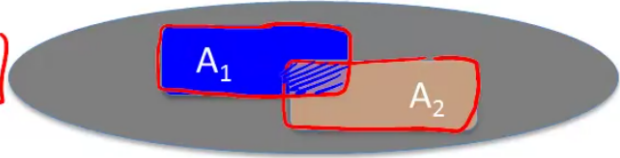
We ask, how likely is it that the output is actually to v ? So formally we say that the probability that X outputs V , is the same as the probability of the event That when we sample a random element in the universe, we fall into the pre image of V under the function X . And again, if this wasn't clear, it's not that important. All you need to know is that a random variable takes values in a particular set V , and in, induces a distribution on that set V . Now there's a particularly important random variable called a uniform random variable. And it's basically defined as you would expect. So let's say that U is some fine [set For example the set of all N bit binary strings, and we're gonna denote a random variable R that's basically sample's uniformly from the set U by this little funny arrow with a little R on top of it. In this, again the notes that the random variable R is literally a uniform random variable sampled over the set U . So in symbols what's this means is that for all elements A in the universe, the probability that R is equal to A is simply R one over U . And if you want to stick to the formal definition of a, of a uniform variable, it's not actually that important. But I would just say that formally the uniform random variable is an identity function namely R .

The union bound

- For events A_1 and A_2 $\subseteq U$

$$\Pr[A_1 \cup A_2] \leq \Pr[A_1] + \Pr[A_2]$$

$$A_1 \cap A_2 = \emptyset \Rightarrow \Pr[A_1 \cup A_2] = \Pr[A_1] + \Pr[A_2]$$



Example:

$$A_1 = \{ \text{all } x \text{ in } \{0,1\}^n \text{ s.t. } \text{lsb}_2(x) = 11 \} ; \quad A_2 = \{ \text{all } x \text{ in } \{0,1\}^n \text{ s.t. } \text{msb}_2(x) = 11 \}$$

$$\Pr[\text{lsb}_2(x) = 11 \text{ or } \text{msb}_2(x) = 11] = \Pr[A_1 \cup A_2] \leq \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

Let r be a uniform random variable on $\{0,1\}^2$

Define the random variable $X = r_1 + r_2$

Then $\Pr[X=2] =$

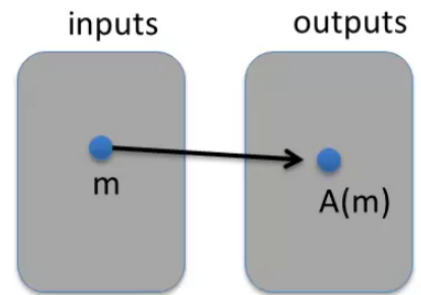
Hint: $\Pr[X=2] = \Pr[r=11]$

is equal to X for all X in the universe So just to see that this is clear, let me ask you a simple puzzle. Suppose we have a uniform random variable over 2-bit strings, so over the set, 01, ten, one and now, let's define a new random variable X to basically sum the first and second bits of R . That is, X simply is the sum of R_1 and R_2 , the first and second bits of R , treating those bits as integers. So, for example, if r happens to be 00, then x will be zero+0, which is zero. So let me ask you, what is the probability that x is equal to two? So it's not difficult to see that the answer is exactly, one-fourth because, basically the only way that x is equal to two is if r happens to be 1,1 but the probability

that r is equal to 1,1 is basically one-fourth because r is uniform over the set of all two bit strings.

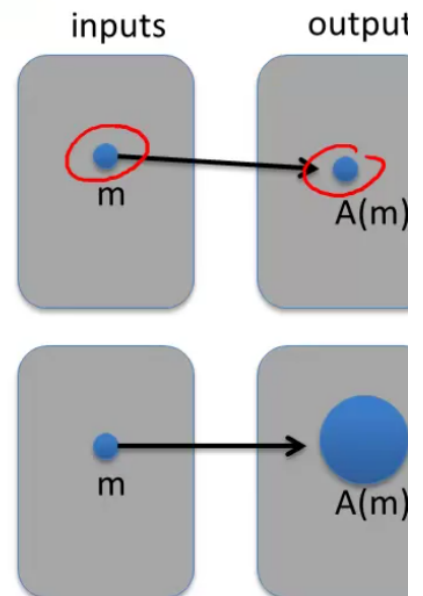
Randomized algorithms

- Deterministic algorithm: $y \leftarrow A(m)$



Randomized algorithms

- Deterministic algorithm: $y \leftarrow A(m)$
- Randomized algorithm
 $y \leftarrow A(m; r)$ where $r \xleftarrow{R} \{0,1\}^n$



The last concept I want to define in this segment is what's called a randomized algorithm. So I'm sure you're all familiar with deterministic algorithms. These are algorithms that basically take a particular, input data, as input, and they always produce the same output, say Y . So if we run the algorithm a hundred times on the same input, we'll always get the same output. So you can think of a deterministic algorithm as a function that given a particular input data, M , will always produce exactly the same output, A of M . A randomized algorithm is a little different, in that, as before, it takes the and as input, but it also has an implicit argument called R , where this R is sampled anew every time the algorithm is run.

Randomized algorithms

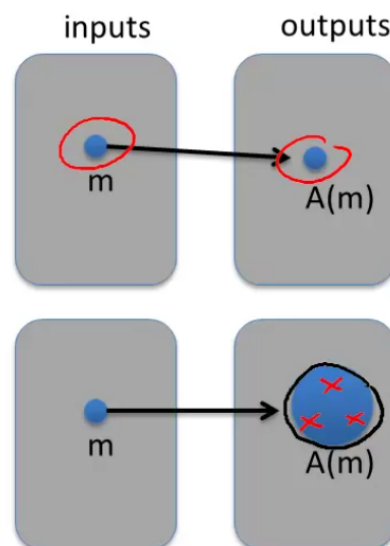
- Deterministic algorithm: $y \leftarrow A(m)$

- Randomized algorithm

$$y \leftarrow A(m; r) \quad \text{where } r \xleftarrow{R} \{0,1\}^n$$

output is a random variable

$$y \xleftarrow{R} A(m)$$



Example: $A(m; k) = F(k, m)$, $y \xleftarrow{R} A(m)$

And in particular this R

is sampled uniformly at random from the set of all N -bit strings, for some arbitrary end. Now what happens is every time we run the algorithm on a particular input M we're gonna get a different output because a different R is generated every time. So the first time we run the algorithm we get one output. The second time we run the algorithm a new R is generated and we get a different output. The third time we run the algorithm a new R is generated and we get a third output and so on. So really the way to think about a randomized algorithm is it's actually defining a random variable. Right? So given a particular input message, M , it's defining a random variable which is, defining a distribution over the set of a \square possible outputs of this algorithm, given the input, M . So the thing to remember is that the output of a randomized algorithm changes every time you run it And in fact, the algorithm defines a distribution and the set of all possible outputs. So let's look at a particular example. So suppose we have a randomized algorithm that takes as input a message M And of course it also takes an implicate input which is this random string that is used to randomize its operation.

So now what the algorithm will do is simply will encrypt the message M using the random string as input. So this basically defines a random variable. This random variable takes values that are encryptions of the message M And really what this random, random variable is it's a distribution over the set of all possible encryptions of the message M under a uniform key. So the main point to remember is that even though the inputs to a randomized algorithm might always be the same every time you run the randomized algorithm you're gonna get a different output. Okay So, that concludes this segment, and we'll see a bit more discrete probability in the next segment.

Over the years many natural cryptographic constructions were found to be insecure. In response, modern cryptography was developed as a rigorous science where constructions are always accompanied by a proof of security. The language used to describe security relies on discrete probability. In this segment and the next, I'll give a brief overview of discrete probability, and I point to this Wiki books article over here for a longer introduction. Discrete probability is always defined over a universe which I'll denote by U and this universe in our case is always going to be a finite set. In fact very commonly our universe is going to be simply the set of all n bit strings which here is denoted by 0^n . So for example the set 0^2 is the set of all two bit strings which happens to be the string zero, zero, zero one, One, zero, and one, one. So there are four elements in this set, and more generally in the set 0^N , there are 2^N elements.

Now a probability distribution over this universe U is simply a function which I'll denote by P , and this function, what it does, is it assigns to every element in the universe a number between zero and one. And this number is what I'll call the weight or the probability of that particular element in the universe. Now there's only one requirement on this function P , and that is that the sum of all the weights, sum up to one.

That is, if I sum the probability of all elements x in the universe, what I end up with is the number one. So let's look at a very simple example looking back to our 2-bit universe. So 00 , ten and eleven And you can consider the following probability distribution Which, for example, assigns to the element 00 , the probability one half.

The elements 01 , we assign the probability $1/8$ th, to ten we assign the probability one quarter and to eleven we assign the probability $1/8$ th. Okay we can see that if we sum up these numbers in fact we get one which means that this probability P is in fact the probability distribution N . Now what these number mean is that if I sample from this probability distribution I'll get the string 00 with probability one half I'll get the string 01 with probability $1/8$ th and so on and so forth.

So now that we understand what a probability distribution is, let's look at two classic examples of probability distributions. The first one is what's called the uniform distribution.**The uniform distribution assigns to every element in the universe, exactly the same weight. I'm gonna use $|U|$ between two bars to denote the size of the universe U . That is the number of elements in the universe, and since we want the sum of all the weights to sum out to one, and we want all these weights to be equal, what this means is that for : Added to Selection. Press to save as a note
(Required)

part2

***. So first

let's do a quick recap of where we are. We said that the discrete probability is always defined over a finite set, which we're gonna denote by U , and typically for us, U is going to be the set of all N bit binary strings, which we denote by 0^N . Now a probability distribution P over this universe U is basically a function that assigns to every element in the universe a weight in the interval zero to one, such that if we sum the weight of all these elements, the sum basically sums up to one

Recap

U: finite set (e.g. $U = \{0,1\}^n$)

Prob. distr. P over U is a function $P: U \rightarrow [0,1]$ s.t. $\sum_{x \in U} P(x) = 1$

. ***Now we have said that subset of the universe is what called an event, and we said that probability of an event is basically the sum of all the weights of all the elements in the event and we said that probability of an event is some real numbers in the interval zero to one And I would remind everyone the basically probability of the entire universe is basically the one by the fact that sum of all the weights sums up to one. Then we define what a random variable is Formally, a random variable is a, is a function from the universe of some other sets But the thing that I want you to remember is that the random variable takes, values in some set v And, in fact, the random variable defines the distribution on this set v

A **random variable** is a function $X: U \rightarrow V$.

Pr / μ

X takes values in V and defines a distribution on V

. So

the next concept we need is what's called independence And I'm gonna very briefly define this If you want to read more about independence, please go ahead and look at the wiki books article.

Recap

U : finite set (e.g. $U = \{0,1\}^n$)

Prob. distr. P over U is a function $P: U \rightarrow [0,1]$ s.t. $\sum_{x \in U} P(x) = 1$

But essentially we say that two events A and B are independent of one another if When you know that event A happens, that tells you nothing about whether event B actually happened or not. Formally, the way we define independence is to say that, the probability of A and B , namely, that both events happened, is actually=to the probability of event A the probability of event B So multiplication, in some sense, the fact that probabilities multiply under conjunction, captures the fact that these events are independent And as I said, if you wanna read more about that, please take a look at the background material. Now the same thing can be said for random variables. So suppose we have two random variables x and y . They take values in some set v . Then we say that these random variables are independent if the probability that $x = a$, and $y = b$ is equal to the product of these two probabilities. Basically what this means is, even if you know that $x = a$, that tells you nothing about the value of y . Okay, that, that's what this multiplication means And again this needs to hold for all A and B in the space of values of these random variables So, just again to jog your memory If you've seen this before, a very quick example. Suppose we look at the, set of, of two bit strings So, zero, zero, zero, one, one zero and, one, one And suppose we choose a random, from this set

Independence

Def: events A and B are **independent** if $\Pr[A \text{ and } B] = \Pr[A] \cdot \Pr[B]$

. Okay so we randomly choose one of these four elements with equal probability. Now let's define two random variables. X is gonna be the least significant bit that was generated, and Y is gonna be the most significant bit generated. So I claim that, these random variables, x and y , are independent of one another. And the way to see that intuitively, is to realize that choosing r uniformly, from the set of four elements is basically the same as flipping a coin An unbiased coin twice. The first bit corresponds to the outcome of the first

flip and the second bit corresponds to the outcome of the second flip. And of course there are four possible outcomes. All four outcomes are equally likely which is why we get the uniform distributions over two bit strings. Now our variables X and Y . Y the independent. Basically if I tell you result of the first flip namely I tell you the least significant bit of R . So I tell you how the first coin you know whether it fell on its head or fell on its tails. That tells you nothing about the result of the second flip. Which is why intuitively, you might expect these random variables to be independent of one another. But formally, we would have to prove that, for, all 01 pairs, the probability of, $x=0$ and $y=0$, $x=1$, $y=1$, and so on. These probabilities multiply.

[RI.png](#)

Independence

Def: events A and B are **independent** if $\Pr[A \text{ and } B] = \Pr[A] \cdot \Pr[B]$

random variables X, Y taking values in V are **independent** if

$$\forall a, b \in V: \Pr[X=a \text{ and } Y=b] = \Pr[X=a] \cdot \Pr[Y=b]$$

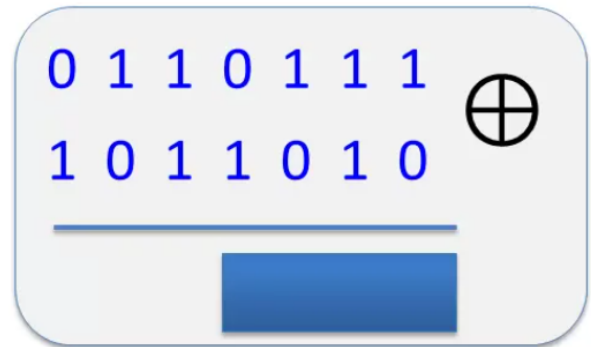
Example: $U = \{0,1\}^2 = \{00, 01, 10, 11\}$ and $r \xleftarrow{R} U$

Define r.v. X and Y as: $X = \text{lsb}(r)$, $Y = \text{msb}(r)$

Let's just do it for one of these pairs. So let's look at the probability that x is equal to zero, and y is equal to zero. Well, you see that the probability that x is equal to zero and y is equal to zero is basically the probability that r is equal to zero, zero. And what's the probability that r is equal to zero, zero? Well, by the uniform distribution, that's basically equal to one-fourth. What is one over the size of the set which is one fourth in this case. And well low and behold that's in fact these probabilities multiply. Because again the probability that X is equal to zero. The probability that the least significant bit of R is equal to zero. This probability is exactly one half because there are exactly two elements that have their least significant bit equal to zero. Two out of four elements gives you a probability of one half. And similarly the probability that Y is equal to zero is also one half so in fact the probability multiplies.

Review: XOR

XOR of two strings in $\{0,1\}^n$ is their bit-wise addition mod 2



Okay, so that's, this concept of independence And the reason I wanted to show you that is because we're gonna look at an, an important property of XOR that we're gonna use again and again. So before we talk about XOR, let me just do a very quick review of what XOR is. So, of course, XOR of two bits means the addition of those bits, modular two. So just too kind of, make sure everybody's on the same page If we have two bits, so 0001, ten and eleven. Their XOR or the truth table of the XOR is basically just the addition modular two. As you can see, one+1 is= to two, modular two. That's=to zero. So this is the truth table for [inaudible] XOR And I'm always going to denote XOR by the circle with a + inside And then when I wanna apply XOR to bit strings, I apply the, addition modular two operation, bitwise. So, for example, the XOR of these two strings would be, 110, and I guess I'll let you fill out the rest of the XORs, just to make sure we're all on the same page. So of course is comes out to one, one zero one. Now, we're gonna be doing a lot of XORing in this class. In fact, there's a classical joke that the only think cryptographers know how to do is just XOR things together But I want to explain to you why we see XOR so frequently in cryptography. Basically, XOR has a very important property, and the property is the following. Suppose we have a random variable y . That's distributed arbitrarily over 01 to the n . So we know nothing about the distribution of y But now, suppose we have an independent random variable that happens to be uniformly distributed also over 01 to the n . So it's very important that x is uniform. N 's independent of y . So now let's define the random variable which is the XOR of x and y .

An important property of XOR

Thm: Y a rand. var. over $\{0,1\}^n$, X an indep. uniform var. on $\{0,1\}^n$

Then $Z := Y \oplus X$ is uniform var. on $\{0,1\}^n$

Proof: (for $n=1$)

$\Pr[Z=0] =$

Y	Pr	Y	X	Pr
0	p_0	0	0	$p_0/2$
1	p_1	0	1	$p_0/2$
		1	0	$p_1/2$
		1	1	$p_1/2$

Then I claim that no matter what distribution y started with, this z is always going to be a uniform, random variable. So in other words if I take an arbitrarily malicious distribution and I XOR with independent uniform random variable what I end up with is a uniform random variable. Okay and this again is kind of a key property that makes x or very useful for crypto. So this is actually a very simple factor to prove, let's just go ahead and do it, let's just prove it for one bit so for $n = \text{one}$. Okay, so the way we'll do it is we'll basically write out the probability distributions for the various random variables. So let's see, for the random variable y . Well, the random variable can be either zero or one. And let's say that P_0 is the probability that it's = to zero, and P_1 is the probability that it's =to one. Okay, so that's one of our tables. Similarly, we're gonna have a table for the variable x . Well, the variable x is much easier. That's a uniform random variable. So the probability that it's=to zero is exactly one half The probability that's it's=to one is also exactly one half. Now, let's write out the probabilities for the joint distribution. In, in other words, let's see what the probability, is for the various, joint values of y and x . In other words, how likely is, it that y is zero, and x is zero. Y is zero, and x is one. Y is one and x is zero, and eleven. Well, so what we do is, basically, because we assume the variables are independent, all we have to do is multiply the probabilities. So The probability that y is equal to zero is p_0 . Probability that x is equal to zero is one-half. So the proximity that, we get 00 as exactly p_0 over two. Similarly for zero one we'll get p_0 over two, for one zero we'll get p_1 over two And for 1-1, again, the probability that y is=to one, and x is=to one, Well, that's P_1 the probability that x is=to one, which is a half, so it's P_1 over two. Okay? So those are the four, probabilities for the various options for x and y .

Proof: (for $n=1$)

$$\Pr[\underline{Z=0}] =$$

$$\Pr[(x,y)=(0,0) \text{ or } (x,y)=(1,1)] = \\ = \Pr[(x,y)=(0,0)] + \Pr[(x,y)=(1,1)] = \frac{p_0}{2}$$

Y	p_r		Y	X	p_r
0	p_0		0	0	$p_0/2$
1	p_1		0	1	$p_0/2$
			1	0	$p_1/2$
			1	1	$p_1/2$

So now, let's

analyze, what is the probability that z is equal to zero? Well, the probability that z is equal to zero is basically the same as the probability that, let's write it this way, xy is equal to 00. Or xy is equal to 11. Those are the two possible cases that z is equal to zero. Because z is the XOR of x and y . Now, these events are disjoint, so, this expression can simply be written as the sum of the two expressions given above. So basically, it's the probability that xy is equal to 00, plus the probability that xy is equal to 11. So now we can simply look up these probabilities in our table. So the probability that xy is equal to 00 is simply $p_0/2$, and the probability that xy is equal to 11, one is simply $p_1/2$. So we get $p_0/2 + p_1/2$. But what do we know, what do we know about p_0 and p_1 ? Well, it's a probability distribution. Therefore, $p_0 + p_1 = 1$. And therefore, this fraction here must be equal to a half. $p_0 + p_1$ is equal to one. So therefore, the sum of these two terms must be a half. And we're done. Basically, we proved that the probability that z is equal to zero is one half, therefore the probability that z is equal to one is also one half. Therefore z is a uniform random variable. So the simple theorem is the main reason why XOR is so useful in cryptography.

English

The birthday paradox

Let $r_1, \dots, r_n \in U$ be indep. identically distributed random vars.

Thm: when $n = 1.2 \times |U|^{1/2}$ then $\Pr[\exists i \neq j: r_i = r_j] \geq \frac{1}{2}$

notation: $|U|$ is the size of U

The last thing I

wanna show you about discrete probability is what's called the birthday paradox. And I'm gonna do it really fast here. Because we're gonna come back later on, and talk

about this in more detail. So, the birthday paradox says the following
 suppose I choose n random variables in our universe u . Okay And it just so happens
 that these variables are independent of one another. They actually don't have to
 be uniform. All we need to assume is that they're distributed in the same way. The
 most important property though is that they're independent of one another. So the
 theorem says that if you choose roughly the square root of the size of u elements,
 we're kind of this one point two here, it doesn't really matter. But if you choose
 square root of the size of u elements, then basically there's a good chance that
 there are two elements that are the same. In other words, if you sample about square
 root a few times, then it's likely that two of your samples. [inaudible] will be
 equal to each other. And by the way, I should point out that this inverted e ,
 just means exists. So there exists in [inaudible] i and j such that r_i is equal
 to r_j . So here's a concrete example. We'll actually see many, many times.
 Suppose our universe consist of all strings of length of one hundred and
 twenty eight bits. So the size of you is gigantic it's actually two to the hundred
 and twenty eight. It's a very, very large set But it so happens if you sample say
 around two the sixty four times from the set.

The birthday paradox

Let $r_1, \dots, r_n \in U$ be indep. identically distributed random vars.

Thm: when $n = 1.2 \times |U|^{1/2}$ then $\Pr[\exists i \neq j: r_i = r_j] \geq \frac{1}{2}$
 exists

notation: $|U|$ is the size of U

Example: Let $U = \{0,1\}^{128}$

$$|U| = 2^{128}$$

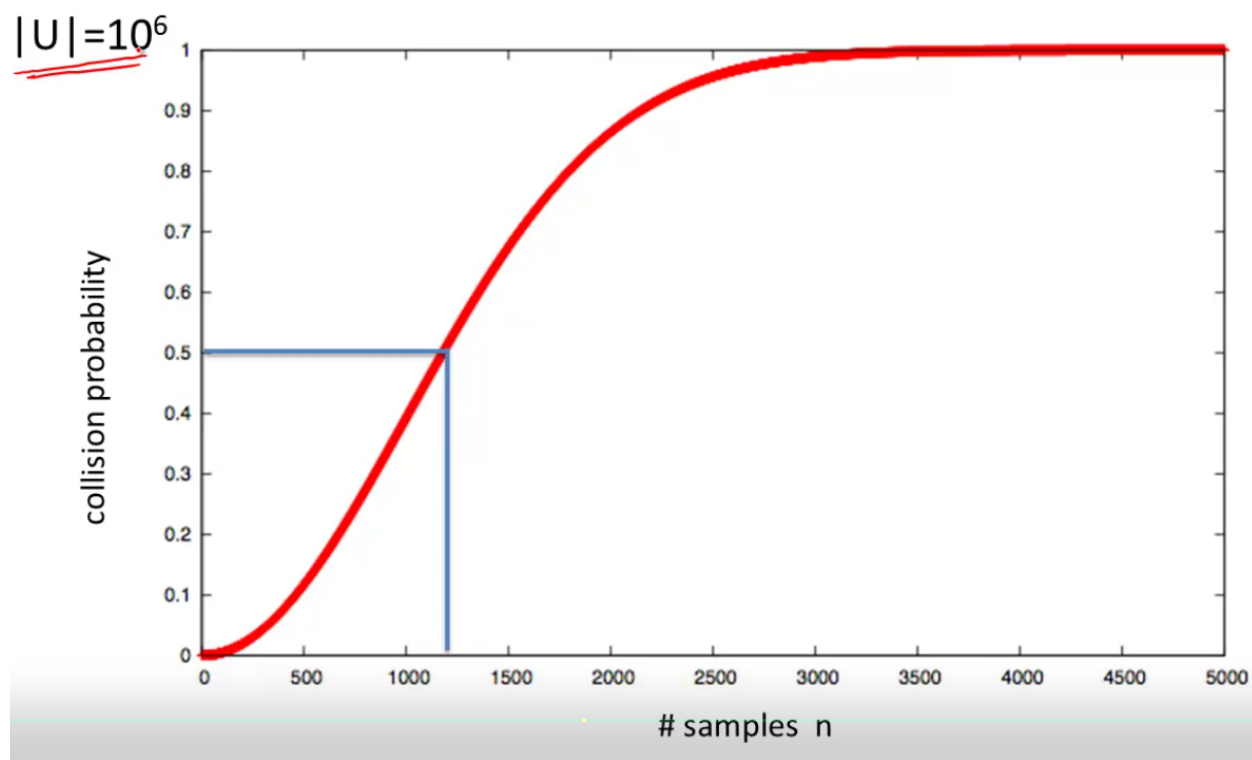
After sampling about 2^{64} random messages from U ,
 some two sampled messages will likely be the same

**This is about the square root of U

then is very likely that two of your sample messages will actually be the same.

So why is, this called the paradox? Well, traditionally it's described in terms of
 people's birthdays. So you would think that each of these samples would be
 someone's birthday And so the question is how many people need to get together so
 that there are two people that have the same birthday? So, just as a simple cal
 culation you can see there are 365 days in the year, so you would need about 1.2
 times the square root of 365 people until the probability that two of them have the

same birthday is more than a half. This if I'm not mistaken is about 24, which means that if 24 random people get together in a room it's very likely that two of them will actually have the same birthday. This is why it's called a paradox, because 24 supposedly is a smaller number than you would expect. Interestingly, people's birthdays are not actually uniform across all 365 days of the year. There's actually a bias towards December. >> But, I guess that's not, that's not relative to the discussion here. >> The last thing I wanted to do is just show you the birthday paradox a bit more concretely. So, suppose we have a universe of about a million samples, then you can see that when we sample roughly 1200 times, the probability that we get, we sample the same number, the same element twice is roughly half But the probability of sampling the same number twice actually converges very quickly to one. You can see that if we about 2200 items, then the probability that two of those items are the same, already is 90 percent and You know, 3000 then it's basically one.



So this conversion is very quickly to one as soon as he goes beyond the square root of the size of the universe. So we're gonna come back and study the birthday paradox in more detail later on, but I just for now, wanted you to know what it is. So that's the end of this segment, and then in the next segment we'll start with our first example of encryption systems. [probability, and I want to remind everyone that if you wanna read more about this,; Added to Selection. Press to save as a note (Required)