

The following content is provided under a Creative Commons license. Your support will help MIT OpenCourseWare continue to offer high quality educational resources for free. To make a donation or view additional materials from hundreds of MIT courses, visit MIT OpenCourseWare at [ocw.mit.edu](http://ocw.mit.edu).

**PROFESSOR:** OK, let's get started. Last week we talked about random variables. And this week we're going to talk about their expected value. The expected value of a random variable comes up in all sorts of applications. We're going to spend a whole week talking about it and some of next week talking about variations from the expected value. And it's probably one of the best tools you have for working with probability problems in practice.

So let's write down the definition. The expected value also has other names. It's also known as the average or the mean of a random variable. Random variable  $R$  over a probability space  $S$  is denoted by a lot of ways. We're going to use  $E_x$  to denote it,  $E_x$  of  $R$ .

And it's the sum over all possible outcomes in the sample space of the value of the random variable on that outcome times the probability of that outcome. In other words, the expected value of a random variable is just a weighted average of all possible values of the random variable where the weight is the probability of that happening.

For example, suppose we roll a fair six sided die. So the numbers that come up are 1 to 6. And we let  $R$  be the random variable denoting the outcome, 1 through 6. So the expected value of the role, we can easily compute from the definition. Well, there's a  $1/6$  chance that it comes out a 1. The next outcome would be a 2. That happens with  $1/6$  chance. 3, 4, 5, and six all happen with a  $1/6$  chance. If I sum up 1, 2, 3, 4 up to 6, I get 6 times  $7/2$  times the  $1/6$  is just  $7/2$  or 3 and  $1/2$ .

So expected value when I roll a die, a fair six sided die, is just 3 and  $1/2$ . And as you can see from this example, the expected value doesn't have to be attainable by one of the outcomes. You don't get a 3 and  $1/2$  when you roll a die. But that's the average value.

Now the expected value, the expectation, the average, the mean, they're all the same thing. They're all defined based on a random variable. They all mean the same thing. They are all different than the median. That's something that's different. So let me define the median.

Basically it is the outcome which splits the probabilities in half. In other words, you've got a 50% chance of being bigger than the median and a 50% chance of being smaller. Now precisely you can define it this way. The median of a random variable  $R$  is the value in the range of  $R$  such that the probability that the random variable is less than this median point is at most  $1/2$ . And the probability that random variable is bigger than the median is strictly less than  $1/2$ .

Now, some texts do it differently. They swap this less than or equal to with this one. And it could give you a different answer if you do that. And I think in the text, actually, we screwed it up. So we have to get that corrected. I think we might have put a less than or equal here, which is not the right definition. So this is the right definition that we'll use.

So using this definition, what is the median of the random variable that corresponds to rolling a die? What's the median when I roll a six sided fair die? Not  $3$  and  $1/2$ , because it's got to be one of the values. The median does have to be in the realm of the random variable. It has to be one of the attainable values. What's the median value when I roll a die?

**AUDIENCE:** Four.

**PROFESSOR:** Four. Let's try that out. If I plug in  $4$  here, the probability I'm less than  $4$  is  $1/2$ . Could be  $1$ ,  $2$ , or  $3$ . The probability of greater than  $4$  is  $5$  or  $6$  is  $1/3$ . That's less than  $1/2$ . So  $4$  works.  $3$  doesn't work, because the probability I'm bigger than  $3$  is  $1/2$ . So it doesn't work. And the other definition sometimes people use would turn out that three is the median.

Now, we're not going to spend any more time talking about the median. What's really important in probability is the mean. And that's because you can do a whole lot more with it in applications. Any question about the definitions so far?

All right, so now we're going to do a more interesting example. We're going to play a gambling game and we're going to analyze the expected winnings, the expected return. And this is a simple three person game that you can see played in bars and informally. But to play it I need a volunteer from the class, somebody who wants to play. You've done it. Who hasn't done it? Have you done it? You haven't played. Do you have money in your pocket? You got to borrow some money. If you don't have any money, you got to borrow some.

**AUDIENCE:** How much money?

**PROFESSOR:** Oh, I don't know,  $\$5$  or  $\$10$ . Got it? All right, come on down. you're going to play against a

couple of TAs. Maybe a couple of you will want to play. I'll get Nick and Martyna to play here.

Now, this is a very simple game. In each round, each of the players is going to wager \$2. So I can loan you guys some money here, I guess. So you have \$2? All right, I have to loan you money. I better count this out.

**AUDIENCE:** I think I can find the Zimbabwe dollars I had here somewhere.

**PROFESSOR:** I don't know if I'm taking those. Here, you can split that up. I loaned you guys \$10 there. You got \$2. Put it on the table. Wager your \$2. Now what we're going to do is they're each going to guess the results of a coin toss. We'll put the pot together here, mix it up. You only have to do two. You can save those for now. Oh, that's your two.

All right, so we have \$6 in the pot. Now, each of you is going to guess the outcome of a coin. So here's the guesses. Heads or tails, that's for you. Heads or tails, the guess for you. And you get a heads or a tails. So instead of writing it down, you're going to sort of hide those, maybe get some advice from the class, and you're going to pick one and put it on the table here as your guess.

And then one of you out there is going to flip a coin and announce the result. And then we're going to reveal their guesses and the winners split the money. And if there's no winners, well, they take their \$2 back and they split the pot.

So this is a very fair game. They make their choices, you guys toss the coin, and the winners share of the pot. If there's one, winner they get \$6, which is a profit of \$4. Two winners, they each get three, profit of one, the other one's out two bucks. Nobody wins, they get their money back, profit 0. They all win, they get their money back, profit 0.

Is the game clear? Who's got a coin out there? You've got a coin? Very good. Now you've got to guess the coin. Don't show each other your guesses. And put your guesses on the table here. You get to pick heads or tails. Put it on the table.

[LAUGHS]

This is not a hard game. You got a heads and a tails, right? Well, yeah, did you put a heads or a tails in there? Yeah, there's one heads, one tails. No, that's blank. You don't win that way. There you go. That's the tails.

**AUDIENCE:** I thought you meant--

**PROFESSOR:** Heads. So if you want heads, you go like that.

**AUDIENCE:** Yeah. I thought you meant do it the other way.

**PROFESSOR:** There you go.

**AUDIENCE:** All right, I'll have to make these indistinguishable now.

**PROFESSOR:** Well, yeah. One of them is all ripped up now. They already guessed, so they've already done their guess. So you pick one and put it there.

**AUDIENCE:** This is what I was originally going to guess anyway.

**PROFESSOR:** So that's good. And you now flip the coin and tell us what it came out.

**AUDIENCE:** Heads.

**PROFESSOR:** Heads. All right, let's reveal your choices. All right, so there's one heads. Martyna takes all the money. Just bad luck there, right? We're going to play again. We're going to play a few times. But let's start recording what happened here. So the coin came out heads. OK, remind me your name?

**AUDIENCE:** Adam.

**PROFESSOR:** Adam gassed tails. And Martyna. Is it Y? Martyna guessed heads and Nick, he also loses at tails. So in this case Adam is out \$2. Martyna got a profit of \$4. And Nick is also down \$2.

All right, let's try it again. So again, pick another heads or tails. Adam's busily working out probabilities here. Well you got to put your money on the table here now. You got \$2. That's three. Here you go.

[LAUGHS]

There we go. Two more dollars. You can make change if you want. There we go. All right, make your choices. You got one? As long as you guys don't show him your choices, we should be good here, I think. And you got to pick one here. Very good. Can we have a coin toss please?

**AUDIENCE:** Heads.

**PROFESSOR:** Heads again. What do we got? All right, good job. OK, Adam got heads and Martyna's on a roll here. So we got to split this up. Can you make change for him here? He gets \$3. So it was heads. Adam had heads, Martyna has heads, and Nick's in trouble. Nick's off another \$2 of my money. Martyna is up \$1 because she wagered two and got three and Adam is up \$1 here for that one.

All right, let's make another selection. Heads or tails? Put the money up here, \$2. Going to need another loan here, Nick?

**AUDIENCE:** I can throw in my keys.

**PROFESSOR:** No, no.

**AUDIENCE:** The keys to my non-existent car.

**PROFESSOR:** You got money? OK. All right.

**AUDIENCE:** I got a lot of money.

**PROFESSOR:** Oh, very good, that's good. That's what we like to hear here.

**AUDIENCE:** That's what he likes to hear.

**PROFESSOR:** Everybody make a choice. Heads or tails? Now, you all want to be thinking, OK, what's going on here? We're going to figure out expectations in a minute, do the tree method. Is there any catch going on? Not that we would have a catch here.

OK, put a heads or tails down. OK, coin toss please.

**AUDIENCE:** Tails.

**PROFESSOR:** Tails. What do we got? A winner, a winner, and a loser. So we got tails, tails, tails. Nick has lost another \$2. Martyna's cruising and Adam is back to even. Good job.

**AUDIENCE:** Can I cash out now?

**PROFESSOR:** Well, you're only even. You've got to get ahead here. All right, let's do it again. \$2. Everybody got \$2 on the table? I'm going to have to loan. You got money? You got it. OK. See, my hope is Nick loses track. OK, make a choice please. Coin toss, please.

**AUDIENCE:** Tails.

**PROFESSOR:** Tails. What do we got? Heads. Martyna, Nick! Oh, for goodness sake.

**AUDIENCE:** She's psychic, I tell you.

**PROFESSOR:** Martyna wins again. You didn't make a deal with him, did you?

**AUDIENCE:** [INAUDIBLE].

**PROFESSOR:** Nick and Martyna are perfect and Adam is down again. Let's collect your money. She gets it all. Oh, she got four here. Yeah, she made four. Good point. Because she won it all. Very good. All right, let's put the next wager up. Nick, if I were you, I wouldn't do much gambling in life. [LAUGHS] You got any money left Adam?

**AUDIENCE:** I'm going to have a big win.

**PROFESSOR:** There it is. Working on the big win. There we go.

**AUDIENCE:** I'm running out of ones here.

**PROFESSOR:** Martyna can make change for you.

**AUDIENCE:** Thanks, Martyna. Do you have change for a 10?

**PROFESSOR:** All right. There we go. A couple bucks there. And then make your selections. Okey doke. It's a pretty fair game. They put the money in, the winners split the pot. You guys flip the coin after they make their selections. OK, make a selection. Martyna has hers. Adam's ready and Nick is ready. Can we have a coin toss?

**AUDIENCE:** Tails.

**PROFESSOR:** Tails. What do we got? Nick! Congratulations. Nick wins big. All right. Tails, heads, heads, tails. Nick, \$4. He's climbing back. And Martyna lost for the first time \$2. Poor Adam is sinking a little bit here.

All right, one more. So who collects? Nick, do you want to collect this? Start paying off the debts, but save \$2 for the last round. This yours? Yeah, there's a last one. All right, two more bucks. Last round. OK, make your selections. Martyna's ready. Nick is ready. OK, Adam, what do you like? Coin toss.

**AUDIENCE:** Heads.

**PROFESSOR:** Heads. How'd you do? Oh Adam, tough break. Nick again. Nick is the sole winner. So he collects on the heads. All right, so it's plus \$4 for Nick, minus 2 for Martina, minus 2 for Adam. All right, let's see how they did in total. Oh, poor Adam. Minus \$6 here. How did that happen? Martyna is plus \$6. Now we know how it happened. And Nick is even. Even. Yeah, so that's just a tough break for Adam. He came up here and lost \$6.

**AUDIENCE:** Interestingly, there's \$6 on the table. I think I should take that and run.

[LAUGHS]

**PROFESSOR:** Yeah, you probably should. But I need to get paid back from Nick here. Now obviously this is a fair game, right? So how many people think that Adam was just unlucky? A few. How many people think we just screwed him out of \$6? Yeah, you're right. But let's do the analysis to see what happened here. Meanwhile, I probably should give him his \$6.

**AUDIENCE:** [INAUDIBLE].

**PROFESSOR:** And you got my money. All right, so I have a gift certificate, if you would like, for playing the game here. You can have this or you can have the pocket protector. I got to say somebody's last time turned in this for one of these. The nerd pride pocket protector. So what would you like for your memento?

**AUDIENCE:** I'm going to take the box where [? David Chena ?] is standing.

**PROFESSOR:** What?

**AUDIENCE:** As in [INAUDIBLE].

**PROFESSOR:** You get one too for tossing the coin. So we'll pass this up. Here you go. Want to pass that up for our coin tosser? OK, very good. Thanks very much. Just leave the money here. Yeah. Leave my money here. Oh, we took some out of your wallet.

**AUDIENCE:** I got \$5 money from you-- \$5.

**PROFESSOR:** \$5 from me, yeah. All right, so leave me the \$5.

**AUDIENCE:** This is your \$5.

**PROFESSOR:** Oh, I should give them back their \$6. Well, otherwise they could report me, and I'd be in trouble. And it's all on film. That would not be good.

OK, so what we're going to do now is analyze the game. And I'm going to prove to you he was just unlucky, despite the fact that you think I screwed him here. So we're going to analyze Adam's expected winnings playing this game. So let's do that.

So we're going to make the tree, as usual. Now, the first thing we have is Adam's choice. Now, Adam were you doing anything besides sort of randomly picking? Were you trying to psych out the coin?

**AUDIENCE:** Apparently I wouldn't be able to bribe him from all the way down here.

**PROFESSOR:** Yeah, and you did lose most of it. We'll say Adam was playing 50-50. It's a reasonable assumption. And then we have Martyna and then we have Nick and their choices. And let's say they're 50-50 as well, because none of them know what coin is going to be tossed. And I'm not going to draw this bottom half of the tree, but it's totally symmetric. It just gets too deep otherwise. And then we have Nick's choices, heads, tails, heads, tails. And they're 50-50.

And then we have the coin. And I'll start drawing it down here. And the coin can be heads or tails. And we're going to assume it was a fair coin toss back there, because it looked like it was flipping a bunch of times. And you don't look like Persi Diaconis, so we're going to assume you didn't learn how to flip heads always. So those are 50-50. A little tight here. Everything is 50-50.

So the probabilities are all  $1/16$ . I got 16 possible outcomes. 2 by 2 by 2 by 2. They're all equally likely. So all the probabilities are  $1/16$ .

And now let's see the winnings. So we'll take Adam's game. It'll turn out to be negative. Now, in this case if Adam's heads and Martyna and Nick are heads and the coin comes up heads, they all win, they split the pot. So how much is the gain for Adam in that case? 0. Now they all guessed heads, but the coin comes out tails, so they split the pot. What's Adam's gain there? 0, because they just split the pot again. He gets his \$2 back. That's 0.

Now we have a case where Nick is tails, Adam and Martyna were heads, and it comes out heads. So Nick is losing here. The pot is split by Adam and Martyna. What does Adam get as a profit? 1. He gets half the part because he splits with Martyna. That's \$3 he put into, he gets

1, plus 1.

Now it turns out same scenario of guesses, but it comes out tails, so Nick wins everything. What's Adam's status here? Minus 2. He bet \$2, he lost it. Then we have heads for Adam, tails for Martyna, heads for Nick. Coin comes up heads, Adam splits with Nick, he gets a net of \$1. Same scenario guesses, but now it's tails, so Martina wins everything, Adam loses \$2.

Now we go down to here where it's heads for Adam, tails for Martyna and Nick, and heads comes up, so Adam wins the whole thing. What does he get in this case? Plus 4. He wins the whole part of 6 minus 2. So he gets 4 net. And then lastly, in this scenario it comes up tails and Martina and Nick split the pot, Adam loses. Minus 2. And then the same thing is happening down here. Same thing. It's just everything is reversed by symmetry.

Now, they're all equally likely. So what we do to compute the expected gain is take each value times its probability and add them up. So 0 times 1/16 plus 0 times 1/16 plus 1 times 1/16 minus 2 times 1/16 and so forth. So it's easier just to add these up and multiply by 1/16. And when we do, we get 0. We get 0. If I add all these up, I get 0. The same thing for down here. So the expected gain for Adam is 0. And that is a fair game.

What do you think? Do you think we'd go to all that trouble just to play a fair game? Or do you think we're trying to set him up and take his money? Yeah.

**AUDIENCE:** Martyna and Nick are alternating, so the branch with the plus 4 goes away.

**PROFESSOR:** Oh, that's interesting. Look what happened here. Nick and Martina, well, they're opposite. That's opposite. That's opposite. That's opposite. You don't suppose they planned that, do you? And how could that possibly help that they just happened to guess opposite?

**AUDIENCE:** One of them always wins.

**PROFESSOR:** One of them always wins. And what does that mean for poor Adam?

**AUDIENCE:** He never takes the whole pot.

**PROFESSOR:** He never takes the whole pot. Well, all right. Let's see if that changed anything here. So if Nick and Martyna are always opposite, that means some of these branches can't occur. So if Martyna's heads, Nick is tails, this is out.

So this is happening with probability 1 now. So these points are at 1/8 each instead of 1/16.

And these go to 0 because that branch can't happen. Same thing here. When you go tails to Martyna, Nick can't be tails, has to be heads. So these go to  $1/8$ . These go to 0.

Now, you notice this isn't working out so well for Adam because we're putting more weight here where he's got a net negative compared to an even situation. And here he's gone from a positive situation, that's getting wiped out to a net negative.

So let's compute his probability now. And I'll get the same contribution down here. I've got  $1$  minus  $2$  plus  $1$  of  $0$  minus  $2$  is negative  $2$ . I'll get a negative  $2$  down here times  $1/8$ . In this case, I'll have the expected gain for Adam is going to be minus. Let's make sure I do this right. It's going to be  $2$  for the top and bottom times  $1/8$  times these guys.  $1$  minus  $2$  plus  $1$  minus  $2$ . And that equals minus  $2$ . That's minus  $1/2$ .

So in fact, if they come up here and guess differently, who knew? Now the expected gain for Adam is minus  $\$0.50$ . Every time he plays, he expects to lose  $\$0.50$  on his  $\$2$  bet. That's a lousy game for Adam to be playing, even though it seemed very fair. You see what's going on here?

Now, this kind of trick is used in all sorts of gambling games. Maybe you've probably played some of these games and may not have realized maybe somebody was using this trick against you.

For example, how many of you have been in some kind of sports betting pool? It's March Madness, you're betting on the victors in round one. It's a football pool for the weekend. You're going to guess against the spread. Who's going to win? All right, some of you have done that. All right, now everybody puts  $\$1$  into the pool. And the winner is the guy who was the most wins or games picked right. All the winners split the pot.

Now, what this says, by doing the same kind of analysis we just did, is that if you collude with one or two or three other players in the pool to always pick differently on all the games, it's going to give you an edge. Same reasoning. When they pick differently, it gives them an edge and now your expected return is bigger than  $0$ , at the expense of the other players who just go in and are putting their picks, if we assume each one is 50-50.

In fact, a former professor of statistics here at MIT used this idea, a guy named Herman Chernoff, in the 1980s to beat the state lottery. Now, everyone knows that lotteries are the worst game around because everybody puts the money in, the state takes half, and then the

winners split the pot. So it's a horrendous game. Your expected return is half of what you put in. So you expect to lose half your bet, because the state's taking half and splitting the remainder among all the participants.

Now, what Chernoff realized is that people don't bet randomly. They tend to pick the same sets of numbers. It might be a birthday. There's only so many birthdays out there. Might be number of home runs Papi's hit, his batting average, Pabelbon's ERA. Who knows? But they pick a relatively small set of numbers that tend to collapse on.

In fact, if you graph how people tend to pick. And say you're playing pick four, where you pick four numbers. And now you look at the frequency with which the numbers are picked. Very crudely it looks something like this. Every once in a while you get a hot ticket where a lot of people pick that number.

For example, if you're picking four numbers, MIT students might pick 2, 4, 6, 16. They might pick that and so it'd be a big spike for that set of four. Down the street they're probably picking in 1, 2, 3, 4. Something like that and there's a spike there.

If you knew this was the histogram of what people were picking and you knew half the pot was going to get split with the winners, what would you pick? Would you pick these? No, because now you're splitting the pot with 100 people. That's no good. You pick down here. And now when you win, you get half the pot. The state always takes half, but you don't have to split it with anybody else. And that means if you're picking down here, your expected return is positive, even with the state taking half. Because so much of the money is piled up in these things.

And so Chernoff proved that in fact, he didn't know which numbers were popular. He didn't know that. So what did he do? If you don't know where the spikes are but you're trying to avoid them and there's not very many spikes, what would you do to avoid them?

Pick random. Because if you pick random, probably you miss the spike. You're down here. With a number nobody would have thought to pick, if some random number. And he showed that, in fact, if you pick randomly, your expected gain for the lottery at the time was 7%, or \$0.7 on the dollar. So positive even with the state taking half.

Now, shortly after that, you saw the proliferation of these random machines that we create random numbers. Because the state wanted to balance it out and not have this kind of a

scenario. So now a lot of the picks are randomly generated in a lot of the lotteries. These things are not immediately obvious, but become very clear once you know the mathematics behind it. Any questions? Yeah.

**AUDIENCE:** How would a person hit a random number?

**PROFESSOR:** Oh, you go to your favorite computer and do a random number generator. Now, that's not perfectly random. People do get into a science of this where they're certain cosmic rays or whatever hitting the earth and the frequency they try to get random numbers out of it or certain kinds of clocks and stuff like that, the tiny low order bits.

Actually getting really independent random numbers can be challenging. A lot of the things you do with a computer generating random numbers, they're distributed nicely, but they're not mutually independent. And there's whole texts that go into how to do that for mutual independence. Getting something that's fair for one of them is not too hard. Any questions about that?

There's another example. How many people ever participated in a Super Bowl bet? OK, like maybe you're trying to guess the over under on point scored. And the person who guesses close to the total number of points wins the pot. And if there's a tie, it's shared.

So in that kind of situation, some people figured out that the average number of points scored in a Super Bowl is, say, 30. And a lot of the guesses then will cluster around 30 points. Now, if you knew this and you knew a lot of the guesses, because there may be people who cared betting, where would you make, say, this is, I don't know, 40 and this is 20, where would you guess?

You could guess here. That's a good one. Or you could guess here. Not only are you not going to share the pot, which helps you, but in this case you'd actually, because being closest, you'd capture all the scores out here. So that's really good. So these are the best guesses to make for your expected return, assuming everybody is guessing around here because they know that's the median. Yeah?

**AUDIENCE:** But those scores aren't likely to happen.

**PROFESSOR:** They're not likely to happen, but it can outweigh splitting the pot with all the people that guessed here. And you get to cover more bases. So you're right, they're less likely to happen because so many people guessed in here where it's likely. Your expected return is better out

here. You may not be likely to win, but your payoff will be very large when you do.

Now, if you're in a bet or a pool with a bunch of 6042 students and they're all guessing out here, well then you want to go back home here. Then it's better. So it all depends on what that distribution looks like. All right, any more questions about this? Yeah.

**AUDIENCE:** [INAUDIBLE] if you're going to be playing a whole bunch of times, and if you're only playing once, wouldn't you look at the most likely result?

**PROFESSOR:** Say that again now?

**AUDIENCE:** If you're going to be only playing once wouldn't you only be concerned with the result that's most likely to show up?

**PROFESSOR:** Well, it depends on your strategy. If you want the expected gain to be maximized, it doesn't matter how many times you're playing. And that can be very different than maximizing your probability of winning. If you want to maximize your probability of winning, you're going to go right at the center point here.

Because if you look at the probability distribution and say the probability distribution looks like that, then you want to bet here, because that's the maximum chance of winning. But say so many people bet there, you'd split 30 ways. Well, this divided by 30 is smaller than this divided by 1. And so your expected return will be bigger out here. So two different things. Maximizing the probability of winning and maximizing your expected return. Yeah?

**AUDIENCE:** When you say maximizing, do you just do derivatives?

**PROFESSOR:** Yeah, then you could do derivatives and all that kind of stuff. That's right. Once you have the curve, you figure it out. Yeah. Yeah.

**AUDIENCE:** Wouldn't it be better to maximize your expected value versus the probability that you win?

**PROFESSOR:** Yes. In general, you want to maximize the expected return on the basis that probably over life you're doing lots of things. And that overall puts you in a better state. Now, we're going to talk about this some next time in terms of taking high risk bets with high, high payoffs. That can maximize your expected return, but you have a very decent chance of losing a lot.

And so you might not want to go there. And we'll talk about that next time we talk about variance and actually what your choice is. Because that is a fundamental choice you face.

Maximizing chance of winning, maximizing expected return. And of course, tied into that is the risk of losing and what kind of risk you're willing to tolerate.

We're going to do a bunch more examples, but before I do, I want to show you some other equivalent definitions of expectation. So the expected value of a random variable, you can also compute it by summing over all possible values of the random variable.  $x$  being in the range of  $R$ .  $x$  times the probability that  $R$  equals  $x$ .

And let's see why this is true. It follows from the definition. From the definition, we know the expected value is the sum over all sample points of the value of the random variable on the sample point times its probability. And now we can organize this sum by the value that  $R$  takes.

So we're going to split this into a double sum where here we're looking first at  $x$  in the range of  $R$ . And then here we look at sample points for which  $R$  on that point is  $x$ . So this sum is equivalent to this one. Here we're just organizing all the sample points for the same value of  $x$ .

All right, now in the inner sum, I've only got values for which  $R$  equals  $x$ , so I can just replace this with  $x$ . The same as before. Only now I've just put  $x$  instead of  $R$ . Now I can pull the  $x$  out since it's a constant independent of that sum.

Now here I'm summing up the probability of all the sample points for which  $R$  of the sample point is  $x$ . And that is just the probability that  $R$  equals  $x$ . That's the definition of the probability of that event. And so now my answer is some overall values  $x$  in the range of  $x$  times of probability  $R$  equals  $x$ . And that's what I was trying to prove.

OK, any questions about that? Now there are some special cases of this when the random variable is on the natural numbers. So the range of  $R$  is the natural numbers.

So a corollary. If the random variable has a range on the natural numbers, then another way to compute the expected value of  $r$  is to simply sum  $i$  equals 1 to infinity  $i$  times the probability  $R$  equals  $i$ . And the proof, well, it's really just saying the same thing here. I'm just summing over the natural numbers. And the case with 0 doesn't matter because I get 0 times the probability of 0. And that adds nothing to the sum. So I just have to sum over the positive integers in that case.

All right, there's another special case of this, which makes it even easier sometimes to

compute the expected value. If  $R$  is a random variable in the natural numbers, then the expected value of  $R$  is simply the sum  $i$  equals 0 to infinity of the probability the random variable is bigger than  $i$ . Which is the same as summing  $i$  equals 1 to infinity, the probability  $R$  is bigger than or equal to  $i$ . They're the same thing.

For example, the first term here, your probability  $R$  bigger than 0. That's the same as saying  $R$  is bigger than or equal to 1 and so forth. So these are clearly the same. And the difference here is we have  $i$  times probability  $R$  equals  $i$ . Here we have probability  $R$  is bigger than  $i$ , with no  $i$  out in front. So let's see why that's true.

Well, we're going to work backwards and evaluate that sum. The sum  $i$  equal 0 to infinity probability  $R$  is bigger than  $i$ . Well, that's the probability  $R$  is bigger than 0 plus the probability  $R$  is bigger than 1 plus probability  $R$  is bigger than 2, and so forth out to infinity. Adding those up.

And now I can write this one out. That's the probability  $R$  equals 1 plus the probability  $R$  equals 2 plus the probability  $R$  equals 3 and so forth. Probability are bigger than 1 equals, well,  $R$  could be 2,  $R$  could be 3, and so forth.

$R$  bigger than 2. Well,  $R$  could be 3, 4, and so forth. And so now if I add all these up, well, I get 1 times probability  $R$  equals 1. Two of these guys. Three of these guys. And you can see the pattern here. Before the next and so forth. And so we've shown that this value equals that value, which is by the corollary just the expected value of  $R$ . So by the corollary, we know that the expected value of  $R$  equals the sum up there. And that's the proof of the theorem.

Sometimes it's easier to use that top expression there. And as a good example, it gives a really easy way to compute the mean time to failure in a system. So let's do that.

Suppose you have a system and it fails with probability  $p$  at each step. And let's assume that the failures are mutually independent. So if the system has been going for  $t$  steps, it still will fail on step  $t$  plus 1 with probability  $p$ , no matter what's happened before.

And the question is, what's the expected number of steps before the system fails? How long is it going to live before you get a crash? And we're going to let  $R$  be that random variable. It will be the step when failure occurs, first failure.

We want to know what's the expected value of  $R$ . Mean time to failure. And we're going to use that top formula up there. Makes it a lot simpler to do that than the other definitions. Now, the

probability that  $R$  is bigger than  $i$  is the same as the probability it did not fail in the first  $i$  steps. So this equals the probability of no failure in the first  $i$  steps.

Because this event,  $R$  bigger than  $i$  means that the first failure was not in the first steps, so the system was fine for the first  $i$  steps. And because of mutual independence on when failure occurs, this is simply the product that we're OK, no failure in the first step times the probability. We're OK in the second step and so forth up to the  $i$ th step. We're OK in  $i$ th step.

And this is where we're using mutual independence, because I've gone from a situation where we're OK in the first  $i$  steps. The probability of that is the product of the probabilities. We're OK in each step individually. So that needed independents.

Well, what is the probability we're OK in the first step? What's the probability of no failure on step one?  $1 - p$ , because  $p$  is the probability we did fail. So we're OK with probability  $1 - p$ . What is the probability we're OK in the second step?  $1 - p$  and so forth for all  $i$  steps. So this is  $(1 - p)^i$ .

And it's usually simpler to write that as  $\alpha^i$  where  $\alpha = 1 - p$ . Because what we've got here now is the sum. Expected value of  $R$  is a sum,  $i$  equals 0 to infinity of that probability, which is just  $\alpha^i$ . And we all know what that sum is, right? What's that sum?  $\frac{1}{1 - \alpha}$ .

And that plug back in the  $\alpha$  to be  $1 - p$  is  $\frac{1}{1 - (1 - p)}$ . And that's very simple. That's  $\frac{1}{p}$ . So the expected time to fail, the expected step of when you're going to fail, is  $\frac{1}{p}$  where  $p$  is the failure probability.

So for example, if you have a 1% chance of failing on any step, your mean time to failure is what? 100.  $\frac{1}{0.01}$ . So very easy to compute mean time to failure. Any questions about that? Of course, you can do it with the other definitions, but the calculations are a little more painful. Yeah.

**AUDIENCE:** Why are we summing it? Is it like an accumulative solution basically?

**PROFESSOR:** Why are we summing? Well, that's the definition. The theorem says the expected value of a random variable is the sum of the probability that it's bigger than  $i$ . That's what the theorem says. So we're just plugging into the theorem. And then the theorem was proved based on the corollary, which came from the theorem and then the definition.

So we went through a series of steps. We started with a definition of expected value, which makes sense. Then we got another way to compute it based on that definition. And then a corollary to that and then we use the corollary to prove another way of computing expected values. We went through a bunch of general steps and then we used, basically you could use this as a definition of this point for expected value. And it just says you sum those things up and you get the answer. Any other questions?

OK, there's a variation on this problem that you see all the time in sort of trick questions or in the popular press sometime, that often confuses people. People sometimes think of it as a paradox, though it's not really one. And the example's something like the following.

Say that a couple, they're going to have kids. What they really want is a baby girl. They get a boy, fine, but that's not what they're concerned about. They want to have a baby girl. And let's say that each time they have a kid it's 50-50 boy or girl. And let's say that it's mutually independent from one kid to the next, which is not true in practice. There tends to be correlation. But let's assume it's mutually independent from one kid to the next.

Now, if on the first try a couple get a girl, great, they're done, they have one kid, that's it, because they just wanted a girl. If they get a boy, OK, try again. And they keep trying again until they get the girl, even if it's 50 kids. They wait till they get the girl. And the question is, how many baby boys do you expect to get before you have the girl? what's? The expected number of boys to get the girl, then you quit? So let's do that.

So we want to know. We have the following data. The probability of a boy is  $1/2$ . You keep having boys until you get a girl, then you quit. We're going to let  $R$  be the random variable for the number of boys. And everything is mutually independent from one child to the next.

How many people think you expect to have more boys than the one girl? You keep having boys till you get the girl. Most people think this. How many people think you expect to have fewer than one boy? Nobody. How many people think you expect to have an equal number? Expect one boy. Good, OK, so that's the answer. In fact, you expect to have just one boy. And the proof, we sort of just did it here.

Now, in this case, we're going to set it up as a mean time to failure. Same kind of thing. Now, in this case the failure mode, they want the girl, but that's when you stop, so that's the failure mode. And a working step is you have a boy and you keep having the working steps until you

have failure mode and then you stop. And in this case, we're not counting the step when you stop.

So we know from that that the expected value of  $R$  is  $1/p$ , which is going to be  $1/2$ . That's the number of children you have. And that equals the mean time to failure. Minus the girl. Because you count the girl as one of the children. And that's going to be the expected number of boys. So number of children you're expected to have minus the girl. This is  $1/2$ . And that is  $2 - 1 = 1$ . So you expect to have one boy before you get the girl. Any questions on that?

OK, how about this? Some couples want to have at least one of each sex. They want to have at least one boy and one girl. So they keep having children until they get one of each and then they stop. How many children do they expect to have? Somebody said two. That's a minimum number. So it's not likely to be the expected number. Three. Well, OK, who said three? Why do you think three?

**AUDIENCE:** Because you have to have at least the first two, so it can be greater than one. I mean, one probability is greater than one child, one probability is greater than two children. And after that it's halves which add up to one.

**PROFESSOR:** Yeah, that's right. That's a good proof. Very good. That'll work. Yeah, yeah.

**AUDIENCE:** I have a question about what you were doing before. If you were to switch your expectations for a girl and boy, wouldn't it come out to the same number but wouldn't it kind of contradict itself?

**PROFESSOR:** No, if you stopped as soon as you had a boy, you'd expect to have one girl before you got a boy. Totally symmetric. In this case, we're stopping when we get the girl, so we expect to have one boy. You might have none, you might have one, you might have two. and as he mentions, if you put the probabilities in there, they will work out the right way, but we just did it simpler by the mean time to failure.

And in fact, there's also another proof, what you're doing. Another way to think about how many kids you have to get one of each. Well you have the first kid. And now you have this problem, because you want one of the other sex. And you keep on trying to get one of the other sex and you expect to have two children to get one of the other sex.

So you have one to start. Whatever it is doesn't matter. And now you expect to have two more

until you hit the other sex. So a total of three is the expected number of children to get one boy and one girl, at least. Any questions about expectation?

OK, let's do another example. This comes up all the time in experimental work. Probably all of you are going to have some example at some point where you're going to do a problem like this. And most all the time, people do it wrong. So let's see an example.

Say that you want to measure the latency of some communications channel. And you want to know what's the average latency. So you set up an experiment. You send a package to the channel, you measure when it started, when it got to the other end, and you record the answer, and then you do it 100 times and you take the average. And you say that's the expected latency of the channel. Very, very typical kind of thing to do, and sometimes OK to do.

So you'd do something like this. You pick a random variable  $D$ , which is going to denote the delay of a packet going through the channel. And there's, of course, some underlying distribution here, which we'll denote by  $f$  of  $x$ . And that's just the probability that  $D$  equals  $x$ . That's the probability distribution function.

And as part of doing your stuff, you notice that if I look at plot  $f$  here, if I look at  $x$  on that axis and  $f$  of  $x$  here, it looks something like this. The chance of getting the observed cases where I got a long delay were small, low probability. And almost all the time I got a short delay. So very typically you'll get a curve that looks something like that in terms of the observations of your experiment.

And say that you do your experiment 100 times and the average latency was 10 milliseconds. And sometimes you want to be really careful, so you do the whole thing again, you do another 100 times. And maybe it's nine milliseconds the next time. And that sort of confirms your belief that your first experiment was valid and that the expected latency on this channel is 10 milliseconds.

Sounds pretty good. But it can be completely wrong. And not just because you're unlucky, but because you're taking a simple method like that. And let me show you an example where it is way off.

All right, say you were just a little more sophisticated about this. And when you did your observations, you tried to figure out what this curve really looks like. And say that you, looking

at the data, concluded that the probability that you have a delay of more than  $i$  milliseconds is  $\frac{1}{i}$ . That looks like this curve, as close as anything.

So from your data, you conclude this. That's a little more sophisticated conclusion, because now you've identified what you believe the distribution is, which is stronger than expectation. How would you go about figuring out the expected value if you had that information? I mean, you could average the 100 sample points. But is there any way, if you assume this, what would you do for the expected value?

Yeah, did I erase the theorem? No, it's over there. We just plug it to the theorem. The expected delay is going to be the sum of those probabilities from 1 to infinity. So we would compute from the theorem the expected delay is  $\sum_{i=1}^{\infty} i \cdot \frac{1}{i^2}$ . Do I want to do the 0 case? Want to be sure I don't get caught up in the 0 case. We'll use the case 1 to infinity. Probability  $D \geq i$ . So let me put  $D \geq i$  here. That equals the sum  $\sum_{i=1}^{\infty} \frac{1}{i}$ .

What's that? What's the sum of  $\frac{1}{i}$ ,  $i$  from 1 to infinity? It's infinite. Remember those harmonic number?  $\sum_{i=1}^n \frac{1}{i}$  gives you about  $\log n$ . Remember that, the book stacking thing?  $\sum_{i=1}^{\infty} \frac{1}{i}$ , that's infinity.

So in fact, your expected latency is infinite. And you just published a paper saying it was 10 milliseconds, maybe nine. So it's very dangerous to just take a bunch of the points, add them up, and average them and say that is the answer. Especially if you have some reason to believe the distribution looks like this and it really is something like that. It could be infinite.

Now in some cases, if your distribution is very well behaved, averaging your sample points is the perfect thing to do. But it helps to know it's not necessarily the case that's a good way to go. Any ideas what went wrong? Yeah.

**AUDIENCE:** I have a question. What would be your probability be over  $i^2$  squared?

**PROFESSOR:** Yeah, what would happen then? If in fact, it was  $\frac{1}{i^2}$ . So you need to do this sum. Here's a good review question for the final. What method do you use to estimate that? Remember the how do we do that? Is that infinite? No. He used the integration bound.

And you'll see what this is pretty small. It'll be, what,  $\frac{1}{2}$ ,  $\frac{1}{2}$ , something like that where we can do the integration method here. So huge difference. This is  $O(1)$ . This is bounded.

Probably less than 2, if we did the integration method. So huge difference here in what the outcome is.

Now, how can it be that I've got something with expected infinite value and I average 100 points and I got 10 milliseconds? Yeah.

**AUDIENCE:** The infinite value comes from the fact that there is a decent probability that delay is going to be huge. However, if you only take a finite number of sample points, then chances are you're not going to get any monstrous delays.

**PROFESSOR:** Exactly. The chance of seeing anything beyond 100 milliseconds is 1 in 100. So I'm probably not going to see. In a sample size of 100, almost surely I won't see something that takes a second, 1,000 milliseconds. But yet it's those rare sample points that are causing that sum to blow up. If I sum that from 1 to 100, I get log of 10. Pretty small. So I get log of 100, which is pretty small.

So what's happening when you do your finite sample is you're missing the big guys which are very rare, but enough to blow up your expectation. Now, you can draw two conclusions from that. One of them is just we did. The other is, well, expectation's the wrong measure. And really what we should be doing is looking at only 1,000 sample points or something like that.

But in practice of using the thing over and over again, eventually you're going to get hit with a whopper. Sometimes you'll see people take data points out when they're really the big ones. They say, oh, well, that was an anomaly. I take that out and then we compute the average. Yeah?

**AUDIENCE:** How much would you pay to play a game where the pair would be the latency of the packet?

**PROFESSOR:** What's that?

**AUDIENCE:** How much do you pay to play a game where the pair would be the latency of the packet?

**PROFESSOR:** All right. That's a big number. If that was my losses here, that's tough. I'd bet anything up against that. To get a payoff of infinity? You'd pay \$1,000 to play that game. Now, you'd want to play it for a long time to get that big payoff, right? But eventually that's what it's going to be. Any questions? People understand the issue here and what to worry about when you all do that someday? Yeah.

**AUDIENCE:** [INAUDIBLE].

**PROFESSOR:** Yeah, now here you don't know for a fact this is it. But you could model and you start seeing these. You fill in the points and you say, it looks like this, let's assume that's what it was, then here's the result. If it looks like it's  $1/i^2$ , you can say, let's assume that's what it is and then you get a different result. But you take the various cases and consider them to do it.

Or you could say, I take the expectation assuming I never get a point bigger than 100. And I limit it that way. And then you're safe at that point and you can get away with it. But to blindly go out there and say here it is, not so reliable.

The expected value does have a lot of useful properties. And the most important is called linearity of expectation. And we'll spend the rest of today and some of next time talking about it. And quite possibly it's one of the reasons people use it so much instead of other things you might think about.

The theorem, and this may be the most important theorem on probability. For any random variables,  $R_1$  and  $R_2$ , on a probability space  $S$ , the expected value of  $R_1 + R_2$  equals the expected value of  $R_1$  plus the expected value of  $R_2$ . Very simple. It's another way of saying expectation is a linear function.

The proof. No, skip the proof here. It's not hard and it's in the text. Follows pretty simply from the definition. There's a generalization for more than two random variables. So corollary. For all  $k$  in the natural numbers and  $k$  random variables  $R_1, R_2, \dots, R_k$  on the sample probability space  $S$ . The expected value of  $R_1 + R_2 + \dots + R_k$  is just the sum of the expected values. And the proof of that is by induction using that result.

And the really important thing about this is that neither of these results needs independence. It is true whether or not the  $R_i$  are independent. Pretty much everything we do in probability to manipulate random variables needs them to be independent. You don't need that here. And that'll make it very powerful.

So let's do an example. Say I roll two fair dice. Six sided dice. Not necessarily independent.  $R_1$  is the outcome on the first die. And  $R_2$  will be the outcome on the second one. And I'm interested in the sum of the dice. So with that  $R$  it'd be  $R_1 + R_2$ . And I want to know the expected value of the sum of the dice when I roll them.

Now, if I didn't use that theorem I'd compute the tree in the sample space. I'd get 36 possible outcomes, take the probability of each. It'd take you a little while to do it, but using linearity of expectation, it's easy. It's expected value of  $R_1$  plus the expected value of  $R_2$ . Each of these we already figured out is  $3\frac{1}{2}$ . And so the answer is 7. So if you roll a pair of dice, whether or not they are independent, the expected sum is seven. Any questions about linearity of expectation? Yeah.

**AUDIENCE:** [INAUDIBLE].

**PROFESSOR:** This one? [INAUDIBLE].

**AUDIENCE:** [INAUDIBLE].

**PROFESSOR:** No, I mean pluses here. Here? I'm computing the sum of the random variables. So this could be a 1, that could be a 10, this could be a 3. So I'm compute the expected value of the sum, just like when I rolled two, dice I'm taking the expected value of the sum of the dice. Any other questions?

Now we're going to do a little trickier problem that uses linearity of expectation. Yeah?

**AUDIENCE:** [INAUDIBLE] sets, like when we're looking at the failure. Couldn't we add the two cases, like  $R_1$  being a girl and  $R_2$  being a boy?

**PROFESSOR:** Well, what would it mean to sum-- so what is the random variable?  $R_1$  is the case when you get a boy. So it's an indicator for getting a boy.  $R_2$  is the indicator for getting a girl.  $R_1$  plus  $R_2$  is by definition 1 then, because you got a boy or a girl. One of them had to happen. And the expected value would be one, But that's a different kind of game. But we are going to start using this in sophisticated ways to make calculations be easier for things like that. so

This problem we call the hat check problem. And the idea here behind this is say that you have  $n$  men at a restaurant having dinner. And when they come into the restaurant, they check their hats in the coat room.

Then something goes wrong in the coat room and the hats are all scrambled up randomly. Let's say a random permutation of the hats. So the men come back to get their hats and they get a random hat coming back. So each man gets a random hat back after dinner.

And the question is, what is the expected number of men to get the right hat back? So we let  $R$

be the random variable that says the number of men to get the right hat, their happen. And we want to know the expected value of  $R$ . Now, from the definition, that's just the sum of all possibilities.  $K$  from 1 to  $n$ .  $K$  times the probability  $R$  equals  $k$ . So using one of the definitions, we could compute the expected value this way. In fact, if you were to be assigned this on a test or on homework, that's probably how you'd start something like that.

And then, well, the next step you'd take would be to figure out what's the probability that exactly  $K$  men get the right hat back. In fact, we actually asked this once before we started doing it in class. And it was really hard if you went down this path. Because if you spent all night with your buddies, you would maybe get to the conclusion that probability  $R$  equals  $K$  is this.  $\frac{1}{K!} \times n \times (n-1) \times \dots \times (n-K+1)$  for  $K$  less than or equal to  $n-2$ . And  $\frac{1}{n!}$  if  $K$  equals  $n-1$  or  $n$ .

Then you would plug that nasty looking thing into there. So multiply by  $K$ . And you'd have to sum it up. And Lord help you. That's just a nightmare to do. You'd have a very hard time getting the answer. If you doubt that, try it. But that would be a natural way to proceed.

But it turns out there is a trivial way to get the answer. And this is a very powerful technique using linearity of expectation. And for sure there will be a problem on the final exam just like this. And so if you go down this path, which is a natural first path, it may take you the rest of the day to solve it. But the method I'm going to show you will take you a couple minutes to do it.

Now, the trick is to use linearity of expectations. So the problem is there's no sum here. So what we need to do is we're going to express  $R$  as the sum of random variables. And the way we're going to do that is as follows. And it's not obvious, but once you see it, it's easy to keep using it.

We let  $R$  be the sum of  $R_1$  plus  $R_2$  plus  $R_n$ . And  $R_i$  is going to tell us the event. This is sort of what you were talking about before with the event of a boy or event of a girl. This is going to be the event of the  $i$ th man gets his right hat back. So it's an indicator random variable. It's 1 if the  $i$ th man gets the right hat. And it's 0 if he doesn't.

So whenever the  $i$ th guy gets his hat back, that counts as one, and now you can see why this sum works.  $R$  is the number of men to get the right hat back. And it basically there's a one counted in here every time a guy gets his hat back, and 0 if he doesn't. So this sum is counting how many men got their right hat back. Non-obvious the first time, gets really simple the fourth

or fifth time. We'll try to do a couple of them today.

All right, now the expected value of  $R$  is easy. It's just by linearity of expectation, expected value of  $R_1$  and so forth plus the expected value of  $R_n$ . The expected value of an indicator random variable is just the probability that it's 1. Right it's 1 times the probability, it's 0 times the probability of 0. That's just the probability that it's 1.

What's the probability that the first man gets the right hat back?  $1/n$ . There's  $n$  hats. He gets a random one. What's the probability that the second man gets his hat back? Not  $1/n$  minus 1. He's coming in whether he's first, second, or last. He gets a random hat.  $1/n$  chance it's his.

What's the probability the last man gets the right hat back?  $1/n$ . Doesn't really matter if he's first or last. Just sort of tricks up a little bit. He's getting a random hat back. So it's  $1/n$ . I got  $n$  of each,  $1/n$ , so what's the expected number of men to get the right hat back? One. Now, the math doesn't get much easier than that.

Now, the amazing thing is we just proved that if we take that mess, stick it in here and sum it up, what answer do we get? 1. That is certainly not obvious, but that is a consequence of everything we've just done. We've just given a probability proof of that fact.

But the nice thing here is there's actually even more powerful. Did I need to assume that it was a random permutation of hats like I would need to assume for this? No independence is needed. In fact, there's all sorts of distributions that will give the same expected value. All I need is that each person gets the right hat back with probability  $1/n$ . In fact, this is an example of a different distribution for which the result is the same.

Say you're at a Chinese restaurant. And you know they have the thing that spins in the middle of the table? Say that you each order an appetizer. There's  $n$  people and they're around a big circular table. Everybody gets their appetizer. Wonton soup, whatever.

And then there's always the joker who spins the thing and spins around and then it stops and now you've got a random appetizer in front of you. In this case, we want to know what's the expected number of people to get the right appetizer back. Not the other guys wonton soup, but yours. That's a different probability space because there's only  $n$  sample points,  $n$  places where the thing could have stopped. Not  $n$  factorial like the hats.

Well, does the analysis change? Exactly the same.  $R_i$  is the indicator variable for the  $i$ th person gets the right appetizer back. Linearity of expectation. The expected value of the indicator variable is just the probability that it's 1. And the probability that any person gets the right appetizer is  $1/n$ .

So the answer is the same. The expected number of people to get the right appetizer back is 1. So totally different probability spaces, exactly the same analysis and answer. OK, so we'll do more of this next time.