

6.041SC Probabilistic Systems Analysis and Applied Probability, Fall 2013 Transcript – Tutorial: A Random Number of Coin Flips

Hey, everyone. Welcome back. Today, we're going to do another fun problem that has to do with a random number of coin flips. So the experiment we're going to run is as follows. We're given a fair six-sided die, and we roll it.

And then we take a fair coin, and we flip it the number of times indicated by the die. That is to say, if I roll a four on my die, then I flip the coin four times. And then we're interested in some statistics regarding the number of heads that show up in our sequence. In particular, we want to compute the expectation and the variance of the number of heads that we see.

So the first step of this problem is to translate the English to the math. So we have to define some notation. I went ahead and did that for us. I defined n to be the outcome of the die roll. Now, since we flip the coin the number of times shown by the die roll, n is equivalently the number of flips that we perform. And n , of course, is a random variable, and I've written its PMF up here.

So P_n of n is just a discrete uniform random variable between 1 and 6, because we're told that the die has six sides and that it's fair. Now, I also defined h to be the number of heads that we see. So that's the quantity of interest. And it turns out that Bernoulli random variables will be very helpful to us in this problem.

So I defined x_i as 1 if the i th flip is heads, and 0 otherwise. And what we're going to do now is, we're going to use these x_i 's to come up with an expression for h . So if you want to count the number of heads, one possible thing you could do is start with 0 and then look at the first coin flip. If it's heads, you add 1 to 0, which I'm going to call your running sum.

If the first flip is tails, you add 0. And similarly, after that, after every trial, if you see heads, you add 1 to your running sum. If you see a tails, you add 0. And in that way, we can precisely compute h . So the mathematical statement of what I just said is that h is equal to x_1 plus x_2 plus x_3 , all the way through x_n .

So now, we are interested in computing $E[h]$, the expectation of h . So your knee jerk reaction might be to say, oh, well, by linearity of expectation, we know that this is an expectation of x_1 , et cetera through the expectation of x_n . But in this case, you would actually be wrong. Don't do that.

And the reason that this is not going to work for us is because we're dealing with a random number of random variables. So each x_i is a random variable. And we have capital n of them. But capital n is a random variable. It denotes the outcome of our die roll.

So we actually cannot just take the sum of these expectations. Instead, we're going to have to condition on n and use iterated expectation. So this is the mathematical statement of what I just said. And the reason why this works is because conditioning on n will take us to the case that we

already know how to deal with, where we have a known number of random variables. And of course, iterated expectations holds, as you saw in lecture.

I will briefly mention here that the formula we're going to derive is derived in the book. And it was probably derived in lecture. So if you want, you can just go to that formula immediately. But I think the derivation of the formula that we need is quick and is helpful. So I'm going to go through it quickly.

Let's do it over here. Plugging in our running sum for h , we get this expression-- x_1 plus x_2 et cetera plus x_n , conditioned on n . And this, of course, is n times the expectation of x_i . So again, I'm going through this quickly, because it's in the book. But this step holds, because each of these x_i 's have the same statistics.

They're all Bernoulli with parameter of $1/2$, because our coin is fair. And so I used x_i to say it doesn't really matter which integer you pick for i , because the expectation of x_i is the same for all i . So this now, the expectation of x_i , this is just a number, it's just some constant, so you can pull it out of the expectation. So you get the expectation of x_i times the expectation of n .

So I gave away the answer to this a second ago. But x_i is just a Bernoulli random variable with parameter of success of $1/2$. And we know already that the expectation of such a random variable is just p , or $1/2$. So this is $1/2$ times expectation of n . And now n we know is a discrete uniform random variable.

And there's a formula that I'm going to use, which hopefully some of you may remember. If you have a discrete uniform random variable that takes on values between a and b -- let's use w -- if you call this random variable w , then we have that the variance of w is equal to b minus a times b minus a plus 2 divided by 12. So that's the variance.

We don't actually need the variance, but we will need this later. And the expectation of w -- actually, let's just do it up here right ahead for this problem. Because we have a discrete uniform random variable, the expectation is just the middle. So you agree hopefully that the middle is right at 3.5 , which is also $7/2$. So this is times $7/2$, which is equal to $7/4$.

So we are done with part of part a. I'm going to write this answer over here, so I can erase. And we're going to do something very similar to compute the variance. To compute the variance, we are going to also condition on n . So we get rid of this source of randomness. And then we're going to use law of total variance, which you've also seen in lecture. And again, the formula for this variance is derived in the book.

So I'm going to go through it quickly. But make sure you understand this derivation, because it exercises a lot of stuff we taught you. So this, just using law of total variance, is the variance of expectation of h given n , plus the expectation of the variance of h given n . And now, plugging in this running sum for h , you get this. It's a mouthful to write.

Bear with me. x_1 through x_n given n -- so I didn't do anything fancy. I just plugged this into here. So this term is similar to what we saw in a previous problem. By linearity of expectation and due

to the fact that all of the x_i 's are distributed in the same way, they have the same expectation, this becomes n times the expectation of x_i . And let's do this term over here.

This term-- well, conditioned on n , this n is known. So we essentially have a finite known sum of independent random variables. We know that the variance of a sum of independent random variables is the sum of the variances.

So this is the variance of x_1 plus the variance of x_2 et cetera, plus the variance of x_n . And furthermore, again, because all of these x_i 's have the same distribution, the variance is the same. So we can actually write this as n times the variance of x_i , where x_i just corresponds to one of the trials. It doesn't matter which one, because they all have the same variance and expectation.

So now, we're almost home free. This is just some scalar. So we can take it out of the variance, but we have to square it. So this becomes expectation of x_i squared times the variance of n . And then this variance is also just a scalar, so we can take it outside.

So then we get variance of x_i times expectation of n . Now, we know that the expectation of x_i is just the probability of success, which is $1/2$. So we have $1/2$ squared, or $1/4$, times the variance of n . So that's where this formula comes in handy.

b is equal to 6, a is equal to 1. So we get that the variance of n is equal to 5 times-- and then 5 plus 2 is 7-- divided by 12. So this is just a formula from the book that you guys hopefully remember. So we get $35/12$. And then the variance of x_i , we know the variance of a Bernoulli random variable is just p times $1 - p$.

So in our case, that's $1/2$ times $1/2$, which is $1/4$. So we get $1/4$. And then the expectation of n , we remember from our previous computation, is just $7/2$. So I will let you guys do this arithmetic on your own time. But the answer comes out to be $77/48$.

So I will go ahead and put our answer over here-- $77/48$ -- so that I can erase. So I want you guys to start thinking about part b while I erase. Essentially, you do the same experiment that we did in part a , except now we use two dice instead of one. So in part b , just to repeat, you now have two dice.

You roll them. You look at the outcome. If you have an outcome of four on one die and six on another die, then you flip the coin 10 times. So it's the same exact experiment. We're interested in the number of heads we want the expectation and the variance. But this step is now a little bit different.

Again, let's approach this by defining some notation first. Now, I want to let n_1 be the outcome of the first die. And then you can let n_2 be the outcome of the second die. And we'll start with just that.

So one way you could approach this problem is say, OK, if n_1 is the outcome of my first die and n_2 is the outcome of my second die, then the number of coin flips that I'm going to make is n_1

plus n_2 . This is the total coin flips. So you could just repeat the same exact math that we did in part a, except everywhere that you see an n , you replace that n with n_1 plus n_2 .

So that will get you to your answer, but it will require slightly more work. We're going to think about this problem slightly differently. So the way we are thinking about it just now, we roll two dice at the same time. We add the results of the die rolls. And then we flip the coin that number of times.

But another way you can think about this is, you roll one die, and then you flip the coin the number of times shown by that die and count the number of heads. And then you take the second die and you roll it. And then you flip the coin that many more times and count the number of heads after that. So you could define h_1 to be number of heads in the first n_1 coin flips.

And you could just let h_2 be the number of heads in the last n_2 coin flips. So hopefully that terminology is not confusing you. Essentially, what I'm saying is, n_1 plus n_2 means you'll have n_1 flips, followed by n_2 flips, for a total of n_1 plus n_2 flips. And then within the first n_1 flips, you can get some number of heads, which we're calling h_1 .

And in the last n_2 flips, you can get some number of heads, which is h_2 . So the total number of heads that we get at the end-- I'm going to call it h^* -- is equal to h_1 plus h_2 . And what part b is really asking us for is the expectation of h^* and the variance of h^* . But here's where something really beautiful happens.

h_1 and h_2 are independent, and they are statistically the same. So the reason why they're independent is because-- well, first of all, all of our coin flips are independent. And they're statistically the same, because the experiment is exactly the same. And everything's independent.

So instead of imagining one person rolling two die and then summing the outcomes and flipping a coin that many times and counting heads, you can imagine one person takes one die and goes into one room. A second person takes a second die and goes into another room. They run their experiments.

Then they report back to a third person the number of heads. And that person adds them together to get h^* . And in that scenario, everything is very clearly independent. So the expectation of h^* -- you actually don't need independence for this part, because linearity of expectation always holds. But you get the expectation of h_1 plus the expectation of h_2 .

And because these guys are statistically equivalent, this is just two times the expectation of h . And the expectation of h we calculated in part a. So this is 2 times $\frac{7}{4}$. Now, for the variance, here's where the independence comes in. I'm actually going to write this somewhere where I don't have to bend over.

So the variance of h^* is equal to the variance of h_1 plus the variance of h_2 by independence. And that's equal to 2 times the variance of h , because they are statistically the same. And the variance of h we computed already. So this is just 2 times $\frac{7}{48}$.

So the succinct answer to part b is that both the mean and the variance double from part A. So hopefully you guys enjoyed this problem. We covered a bunch of things. So we saw how to deal with having a random number of random variables.

Usually we have a fixed number of random variables. In this problem, the number of random variables we were adding together was itself random. So to handle that, we conditioned on n . And to compute expectation, we use iterated expectation.

To compute variance, we used law of total variance. And then in part b, we were just a little bit clever. We thought about how can we reinterpret this experiment to reduce computation. And we realized that part b is essentially two independent trials of part a. So both the mean and the variance should double.

MIT OpenCourseWare
<http://ocw.mit.edu>

6.041SC Probabilistic Systems Analysis and Applied Probability
Fall 2013

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.