

6.041SC Probabilistic Systems Analysis and Applied Probability, Fall 2013 Transcript – Tutorial: A Coin with Random Bias

Hi. In this problem, we're going to be dealing with a variation of the usual coin-flipping problem. But in this case, the bias itself of the coin is going to be random. So you could think of it as, you don't even know what the probability of heads for the coin is.

So as usual, we're still taking one coin and we're flipping it n times. But the difference here is that the bias is because it was random variable Q . And we're told that the expectation of this bias is some μ and that the variance of the bias is some σ^2 , which we're told is positive. And what we're going to be asked is find a bunch of different expectations, covariances, and variances.

And we'll see that this problem gives us some good exercise in a few concepts, a lot of iterated expectations, which, again, tells you that when you take the expectation of a conditional expectation, it's just the expectation of the inner random variable. The covariance of two random variables is just the expectation of the product minus the product of the expectations. Law of total variance is the expectation of a variance, of a conditional variance plus the variance of a conditional expectation. And the last thing, of course, we're dealing with a bunch of Bernoulli random variables, coin flips. So as a reminder, for a Bernoulli random variable, if you know what the bias is, it's some known quantity p , then the expectation of the Bernoulli is just p , and the variance of the Bernoulli is p times $1 - p$.

So let's get started. The problem tells us that we're going to define some random variables. So x_i is going to be a Bernoulli random variable for the i coin flip.

So x_i is going to be 1 if the i coin flip was heads and 0 if it was tails. And one very important thing that the problem states is that conditional on Q , the random bias, so if we know what the random bias is, then all the coin flips are independent. And that's going to be important for us when we calculate all these values.

OK, so the first thing that we need to calculate is the expectation of each of these individual Bernoulli random variables, x_i . So how do we go about calculating what this is? Well, the problem gives us a hint. It tells us to try using the law of iterated expectations. But in order to use it, you need to figure out what you need the condition on.

What this is? What takes place in y ? And in this case, a good candidate for what you condition on would be the bias, the Q that we're unsure about. So let's try doing that and see what we get.

So we write out the law of iterated expectations with Q . So now hopefully, we can simplify it with this inter-conditional expectation is. Well, what is it really? It's saying, given what Q is, what is the expectation of this Bernoulli random variable x_i ?

Well, we know that if we knew what the bias was, then the expectation is just the bias itself. But in this case, the bias is random. But remember a conditional expectation is still a random variable.

And so in this case, this actually just simplifies into Q . So whatever the bias is, the expectation is just equal to the bias. And so that's what it tells us. And this part is easy because we're given that the expectation of q is μ .

And then the problem also defines the random variable x . X is the total number of heads within the n tosses. Or you can think of it as a sum of all these individual x_i Bernoulli random variables. And now, what can we do with this? Well we can remember that linearity of expectations allows us to split up this sum. Expectation of a sum, we could split up into a sum of expectations.

So this is actually just expectation of x_1 plus dot dot dot plus all the way to expectation of x_n . All right. And now, remember that we're flipping the same coin. We don't know what the bias is, but for all the n flips, it's the same coin. And so each of these expectations of x_i should be the same, no matter what x_i is.

And each one of them is μ . We already calculated that earlier. And there's n of them, so the answer would be n times μ .

So let's move on to part B. Part B now asks us to find what the covariance is between x_i and x_j . And we have to be a little bit careful here because there are two different scenarios, one where i and j are different indices, different tosses, and another where i and j are the same. So we have to consider both of these cases separately.

Let's first do the case where x and i are different. So i does not equal j . In this case, we can just apply the formula that we talked about in the beginning. So this covariance is just equal to the expectation of x_i times x_j minus the expectation of x_i times expectation of x_j .

All right, so we actually know what these two are, right? Expectation of x_i is μ . Expectation of x_j is also μ . So this part is just μ squared. But we need to figure out what this expectation of x_i times x_j is.

Well, the expectation of x_i times x_j , we can again use the law of iterated expectations. So let's try conditioning on Q again. And remember we said that this second part is just μ squared.

All right, well, how can we simplify this inner-conditional expectation? Well, we can use the fact that the problem tells us that, conditioned on Q , the tosses are independent. So that means that, conditioned on Q , x_i and x_j are independent.

And remember, when random variables are independent, the expectation of product, you could simplify that to be the product of the expectations. And because we're in the condition world on Q , you have to remember that it's going to be a product of two conditional expectations. So this will be expectation of x_i given Q times expectation of x_j given Q minus μ squared still.

All right, now what is this? Well the expectation of x_i given Q , we already argued earlier here that it should just be Q . And then the same thing for x_j . That should also be Q . So this is just expectation of Q squared minus μ squared.

All right, now if we look at this, what is the expectation of Q squared minus μ squared? Well, remember μ is just, we're told that μ is the expectation of Q . So what we have is the expectation of Q squared minus the quantity expectation of Q squared.

And what is that, exactly? That is just the formula or the definition of what the variance of Q should be. So this is, in fact, exactly equal to the variance of Q , which we're told is σ^2 .

All right, so what we found is that for i not equal to j , the coherence of x_i and x_j is exactly equal to σ^2 . And remember, we're told that σ^2 is positive. So what does that tell us? That tells us that x_i and x_j , or i not equal to j , these two random variables are correlated.

And so, because they're correlated, they can't be independent. Remember, if two intervals are independent, that means they're uncorrelated. But the converse isn't true. But if we do know that two random variables are correlated, that means that they can't be independent.

And now let's finish this by considering the second case. The second case is when i actually does equal j . And in that case, well, the covariance of x_i and x_i is just another way of writing the variance of x_i . So covariance, x_i, x_i , it's just the variance of x_i .

And what is that? That is just the expectation of x_i squared minus expectation of x_i quantity squared. And again, we know what the second term is. The second term is expectation of x_i quantity squared. Expectation of x_i we know from part A is just μ , right? So that's just second term is just μ^2 .

But what is the expectation of x_i squared? Well, we can think about this a little bit more. And you can realize that x_i squared is actually exactly the same thing as just x_i .

And this is just a special case because x_i is a Bernoulli random variable. Because Bernoulli is either 0 or 1. And if it's 0 and you square it, it's still 0. And if it's 1 and you square it, it's still 1.

So squaring it doesn't really doesn't actually change anything. It's exactly the same thing as the original random variable. And so, because this is a Bernoulli random variable, this is exactly just the expectation of x_i .

And we said this part is just μ^2 . So this is just expectation of x_i , which we said was μ . So the answer is just μ minus μ^2 .

OK, so this completes part B. And the answer that we wanted was that in fact, x_i and x_j are in fact not independent. Right.

So let's write down some facts that we'll want to remember. One of them is that expectation of x_i is μ . And we also want to remember what this covariance is.

The covariance of x_i and x_j is equal to σ^2 when i does not equal j . So we'll be using these facts again later. And the variance of x_i is equal to $\mu - \mu^2$.

So now let's move on to the last part, part C, which asks us to calculate the variance of x in two different ways. So the first way we'll do it is using the law of total variance. So the law of total variance will tell us that we can write the variance of x as a sum of two different parts. So the first is variance of x expectation of the variance of x conditioned on something plus the variance of the initial expectation of x conditioned on something. And as you might have guessed, what we're going to condition on is Q .

Let's calculate what these two things are. So let's do the two terms separately. What is the expectation of the conditional variance of x given Q ?

Well, what is-- this, we can write out x . Because x , remember, is just the sum of a bunch of these Bernoulli random variables. And now what we'll do was, well, again, use the important fact that the x 's, we're told, are conditionally independent, conditional on Q .

And because they're independent, remember the variance of a sum is not the sum of the variance. It's only the sum of the variance if the terms in the sum are independent. In this case, they are conditionally independent given Q . So we can in fact split this up and write it as the variance of x_1 given Q plus all the way to the variance of x_n given Q .

And in fact, all these are the same, right? So we just have n copies of the variance of, say, x_1 given Q . Now, what is the variance of x_1 given Q ?

Well, x_1 is just a Bernoulli random variable. But the difference is that for x , we don't know what the bias or what the Q is. Because it's some random bias Q .

But just like we said earlier in part A, when we talked about the expectation of x_1 given Q , this is actually just Q times $1 - Q$. Because if you knew what the bias were, it would be p times $1 - p$. So the bias times $1 - \text{the bias}$.

But you don't know what it is. But if you did, it would just be q . So what we do is we just plug in Q , and you get Q times $1 - Q$.

All right, and now this is expectation of n . I can pull out the n . So it's n times the expectation of $Q - Q^2$, which is just n times expectation $Q - Q^2$, we can use linearity of expectations again, expectation of Q is μ .

And the expectation of Q^2 is, well, we can do that on the side. Expectation of Q^2 is the variance of Q plus expectation of Q quantity squared. So that's just $\sigma^2 + \mu^2$. And so this is just going to be then $n(\mu - \sigma^2 - \mu^2)$.

All right, so that's the first term. Now let's do the second term. The variance the conditional expectation of x given Q . And again, what we can do is we can write x as the sum of all these x_i 's.

And now we can apply linearity of expectations. So we would get n times one of these expectations. And remember, we said earlier the expectation of x_1 given Q is just Q . So it's the variance of n times Q .

And remember now, n is just-- it's not random. It's just some number. So when you pull it out of a variance, you square it. So this is n squared times the variance of Q .

And the variance of Q we're given is σ^2 . So this is n squared times σ^2 . So the final answer is just a combination of these two terms. This one and this one.

So let's write it out. The variance of x , then, is equal to-- we can combine terms a little bit. So the first one, let's take the μ s and we'll put them together. So it's $n\mu - \mu^2$.

And then we have n squared times σ^2 from this term and minus n times σ^2 from this term. So it would be n squared minus n times σ^2 , or n times $n - 1$ times σ^2 . So that is the final answer that we get for the variance of x .

And now, let's try doing it another way. So that's one way of doing it. That's using the law of total expectations and conditioning on Q . Another way of finding the variance of x is to use the formula involving covariances, right? And we can use that because x is actually a sum of multiple random variables x_1 through x_n .

And the formula for this is, you have n variance terms plus all these other ones. Where i is not equal to j , you have the covariance terms. And really, it's just, you can think of it as a double sum of all pairs of x_i and x_j where if i and j happen just to be the same, that it simplifies to be just the variance. Now, so we pulled these n terms out because they are different than these because they have a different value.

And now fortunately, we've already calculated what these values are in part B. So we can just plug them in. All the variances are the same. And there's n of them, so we get n times the variance of each one. The variance of each one we calculated already was $\mu - \mu^2$.

And then, we have all the terms where i is not equal to j . Well, there are actually n squared minus n of them. So because you can take any one of the n 's to be the first to be i , any one of the n to be j . So that gives you n squared pairs.

But then you have to subtract out all the ones where i and j are the same. And there are n of them. So that leaves you with n squared minus n of these pairs where i is not equal to j .

And the coherence for this case where i is not equal to j , we also calculated in part B. That's just σ^2 . All right, and now if we compare these two, we'll see that they are proportionally

exactly the same. So we've use two different methods to calculate the variance, one using this summation and one using the law of total variance.

So what do we learn from this problem? Well, we saw that first of all, in order to find some expectations, it's very useful to use law of iterated expectations. But the trick is to figure out what you should condition on. And that's kind of an art that you learn through more practice.

But one good rule of thumb is, when you have kind of a hierarchy or layers of randomness where one layer of randomness depends on the randomness of the layer above-- so in this case, whether or not you get heads or tails depends on, that's random, but that depends on the randomness on the level above, which was the random bias of the coin itself. So the rule of thumb is, when you want to calculate the expectations for the layer where you're talking about heads or tails, it's useful to condition on the layer above where that is, in this case, the random bias. Because once you condition on the layer above, that makes the next level much simpler. Because you kind of assume that you know what all the previous levels of randomness are, and that helps you calculate what the expectation for this current level. And the rest of the problem was just kind of going through exercises of actually applying the--

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>.