Example question 2: complete work-through of the final example from Lecture 4.

Suppose you are watching a train track where the length of each train is a random variable following the Normal distribution with mean 3,000ft and standard deviation 600ft. Another way to write this is "the lengths of trains are random with distribution $\mathcal{N}(3,000\text{ft}, 600\text{ft})$". **Question**: what is the probability of seeing two (independent) trains in a row that are each greater than 4,000 feet long?

Solution: (Important: Because this solution is meant to be explanatory, it contains a lot of words and sentences that you will not need to write out on a quiz or midterm to receive full credit.)

Let $X_1 =$ the length of the first train and $X_2 =$ the length of the second train (the 1 and 2 are just labels for the variables; they don't say anything about the variables' values). Using these symbols, the question is asking for the probability of the event "$X_1 > 4,000$ and $X_2 > 4,000$". So, our answer should be the number on the other side of this equation:

$$P(X_1 > 4,000 \text{ and } X_2 > 4,000) = ?$$

We can solve for this probability using probability rules and properties of the Normal density. First we note that since the trains are independent,

$$P(X_1 > 4,000 \text{ and } X_2 > 4,000) = P(X_1 > 4,000)P(X_2 > 4,000)$$

by the multiplication rule. So we just need to solve for $P(X_1 > 4,000)$ and $P(X_2 > 4,000)$ individually. One thing to realize instantly is that X_1 and X_2 have the same distribution and density, so

$$P(X_1 > 4,000) = P(X_2 > 4,000)$$

This makes the problem easier: we can just solve for one of them. Let's do that. From the problem we know that X_1 is distributed $\mathcal{N}(3000, 600)$. Define the random variable $Z = \frac{X_1 - 3000}{600}$. Then from the fact in the slides (and in your textbook), Z has the Standard Normal density. The next step then is to write the probability of interest in terms of Z:

$$P(X_1 > 4,000) = P\left(\frac{X_1 - 3000}{600} > \frac{4,000 - 3000}{600}\right) = P(Z > 1.66)$$

Those probabilites are all equal because *the events inside them are the same*. We have *transformed* a Normal probability about X_1 into a Normal probability about Z, which is what we need to get a cumulative proportion. So 1.66 is the z-score, and the area under the curve to the right of $z = 1.66$ is (using the Normal table) 0.0485. Therefore

$$P(X_1 > 4,000) = 0.0485$$

and thus

$$P(X_1 > 4,000 \text{ and } X_2 > 4,000) = 0.0485^2 = 0.00235$$

That's the answer!

If you think about it, the answer kinda makes sense. The statement "trains are random with density $\mathcal{N}(3,000\text{ft}, 600\text{ft})$" is the same as the statement "the full train population is approximately Normal with mean 3,000ft and standard deviation 600ft". One train of length 4,000ft is almost 2 standard deviations above the mean: so seeing one at random is pretty unlikely! And seeing *two in a row* is even *more* unlikely, if they are independent!

Another way to conceptualize the connection between probability and populations is the following. Consider that asking

What is the probability that a random train has length $> 4,000$ft?

is the same thing as asking

What proportion of all trains have length $> 4,000$ft?

Our strategy for finding the probability (using a density) was the same strategy for finding proportions of populations (using a density) in Lecture 1.