Select Reading Question Responses (9/6/2021)

> I am confused with the terms where a trial can be considered an independent Bernoulli trial. What exactly does "independent" mean in the topic of Bernoulli?

Formally, two random variables X and Y are *independent* if, for every possible value x of X and every possible value y of Y, the events $X = x$ and $Y = y$ are independent of each other. In other words, X and Y are independent if

$$P(X = x \text{ and } Y = y) = P(X = x)P(Y = y)$$

for all values x of X and y of Y.

More intuitively, two random variables are independent if knowing the value of one of them gives you no further information about the value of the other random variable. In particular, two Bernoulli trials are independent if knowing the result of one of them gives you no information about the result of the other.

For example, let's say I'm flipping a fair coin and "success" is heads. I flip it once. That's a Bernoulli trial, let's call it X. I flip it again, and that's another Bernoulli trial Y. Even if I know that the first flip landed heads (ie, even if know that $X = 1$), the probability of "success" on the second flip is still 50% — knowing that the first one landed heads doesn't change the probabilities of the second Bernoulli trial. This tells us that X and Y are independent.

> I am curious how we calculate the area under the curve for distributions in order to get a percentile value. For the homework problem I used a specific calculator but is there a way to do this by hand or a strategy for getting a rough estimate.

> Is there a way to solve guided practice 4.15 without a computer program?

There's no quick and precise way of going between z-scores and percentiles for normal distributions "by hand." Before the days of computers, people did a lot of very difficult computations by hand and then compiled the results in a table. Then, when people needed to go between z-scores and percentiles, they just used those tables that told you what z-scores corresponded to what percentiles for the normal distribution. Nowadays, computers can do those difficult computations on the fly, so people just use computers.

But there are *imprecise* way of going betweeen z-scores and percentiles, using the 68-95-99.7 rule. For example, suppose you know the z-score of an observation from a normal random variable is 2. We know that 95% of the data is within 2 standard deviations of the mean, so 5% is in the tails, and 2.5% is within each tail. That means that 95% is in the area that's within 2 standard deviations of the mean, and 2.5% is *below* two standard deviations from the mean, so the percentile corresponding to a z-score of 2 would be 95% + 2.5% =

97.5%. If you had a z-score close to 2 (like 1.9, or so), you could make the same percentile approximation.

> I was wondering if there was a test we could apply to determine if a data set resembled a normal distribution close enough to determine if the principles of the normal distribution could be applied.

There isn't anything very definitive, which makes some sense since "resembling" is a vague notion. But there are a bunch of heuristics you might apply. You might look at a histogram to decide if a distribution looks sufficiently normal. In lab 4, you'll make "Q-Q plots" to determine if things look normal. For binomial random variables, there is the $np, n(1-p) \geqslant 10$ condition for approximating using a normal distirbution.

> In section 4.1, specifically on percentiles, the book provides an example on finding someone's height based on what percentile they belong to within a sample. I just don't understand how you can get an exact number and assign it to an individual that is a member of the percentile. To me it makes sense that a specific height could be a maximum or minimum for that percentile of the sample population, but to be able to draw conclusions about someone's exact height solely based on the percentile they belong to, I just don't get how that's possible.

They're using the fact here that heights are normally distributed (cf. the sentence sandwiched between guided practice 4.11 and 4.12). If you know that someone's height is at the 50th percentile, for example, that means their height is above exactly 50% of people. Since heights are normally distributed, this means that their height must be *exactly the mean* height. You can do similar calculations for other percentiles (but now you'll need the `qnorm` function).

If you didn't know that heights are normally distributed, you're absolutely correct that there would be no way of going from percentile back to the actual height!

> Why does $0! = 1$ instead of 0? I was confused trying to calculate the fraction with $5!/0!$, as it seemed like the denominator would have been zero making it impossible to calculate.

There are two ways of answering this question, I suppose! The first way is just to say "It's a convention that $0! = 1$." The obvious follow-up question you would then ask me is "Why is this the convention, as opposed to something like $0! = 0$ or even $0! = 783442834$ or whatever?" I would then answer that "$0! = 1$ is a *useful* convention, because it makes formulas work the way that you would want for them to work." For example, if you have a binomial random variable X and you want to know what is the probability of observing 0

successes, you would be able to use the formula $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$ with $k = 0$ only if you were using the convention $0! = 1$.

I think that's the easier answer to understand. Unless you're feeling in a deeply philosophical mood, probably that's a convincing enough reason. But if you are feeling deeply philosophical... I think there is a deeper reason that the convention $0! = 1$ should be the useful one.

When n is a positive integer, we define $n!$ to be the product of the positive integers less than or equal to n. When $n = 0$, *there are no positive integers* less than or equal to 0, so it's like we're taking an empty product (ie, a product of nothing).

If I told you to *add* together nothing, it would be fair for you to say that the sum is 0. The underlying reason for this is that 0 is the additive identity: it is the number with the property that $x + 0 = x$ for all numbers x. But here, we're *multiplying* nothing. The *multiplicative* identity isn't 0 — it's 1, since that's the number that has the property that $x \cdot 1 = x$ for all nonzero numbers x. If I interpret $0!$ to mean "the product of nothing," it should be that $0!$ is the multiplicative identity, ie, 1 (by the same logic that allowed me to say that "the sum of nothing is 0.") This is the deepest reason I know for why it's reasonable to expect $0! = 1$ to be a useful convention.

> This is more of a curiosity question, but how did the Bernoulli distribution come to be? The concept seems pretty straight forward, so I guess my question is what did Bernoulli do to discover this distribution?

This is a bit of a tangential answer, because I'm not really a historian of mathematics and I haven't studied Bernoulli's original works in any detail. Roughly what I know is that Bernoulli did a lot of work on understanding probabilistic processes. What we now call the "Bernoulli distribution" models one of the simplest possible types of probabilistic processes, so it's not too surprising in retrospect that someone thining a lot about probabilistic processes would end up describing this process in some fashion. That being said, it is not remotely my intention to minimize Bernoulli's achievement! You might recall that on the first day of class, someone asked me why I liked math (or something like that) and I said something about identifying patterns. To me, a big part of what math *is* involves identifying similarities in lots of different situations and writing down abstractions that encapsulate something important about all of those situations. Flipping a coin once and seeing if it lands heads, sampling one random American and seeing if they speak Spanish at home, seeing if couple's first child has blue eyes or not, etc — on the face of it, these seem like very different probabilistic processes. The way that I would understand Bernoulli's achievement is that he realized that these probabilistic processes, which seem so different superficially, are actually all very similar in some important ways! What we now call the "Bernoulli distribution" is a mathematical abstraction which models all of these probabilistic processes.