MATH 243, LECTURE 12

1. Probability distributions

To be systematic about probabilities in the continuous setting, we have to formalize the notion of distribution.

Definition 1. (1) *A distribution for a continuous random variable can be any non-negative function D where the total area under the function is* 1.
(2) *If X is distributed according to a distribution D then $P(a \leq X \leq b)$ is the area between the line $x = a$, the line $x = b$, the graph of D, and the x-axis.*
(3) *The sample space for such a variable is all x where D is non-zero.*
(4) *If D is equal to either 0 or one other fixed number, we say D is a uniform distribution.*

Example 2. *Revisit the example about numbers between 0 and 3.*

- *Our sample space is the closed interval* $[0, 3]$.
- *Our probability distribution is the function*

$$p(t) = \begin{cases} 1/3 & 0 \leq t \leq 3 \\ 0 & \textit{else} \end{cases}$$

Again, what is the probability that some randomly chosen number is between 1.5 and 2.4? Etc.

Example 3. *Which of the following are distributions?*

- $f(x) = 1$ *for x between -2 and -1, and zero otherwise.*
- $f(x) = 1$ *over the whole real number line.*
- $f(x) = \frac{x}{2}$ *for x between 0 and 2 and zero otherwise.*
- $f(x) = \frac{x}{2}$ *for x between -1 and 1 and zero otherwise.*

Example 4. *Suppose that X is distributed according to*

$$D = \begin{cases} \frac{1}{12} & 1 \leq x \leq 3 \\ \frac{1}{2} & 3 \leq x \leq 4 \\ \frac{1}{9} & 4 \leq x \leq 7 \\ 0 & \textit{otherwise} \end{cases}$$

- *What is the state space?*
- *What is $P(2 \leq X \leq 3)$?*
- *Which is more likely, that X is between 3 and 4 or that X is between 5 and 7?*

Compare the answer to the last question to the answer you would get if X were uniformly distributed.

One way to get a distribution is to take the histogram of a data set and divide by the total number of individuals so that the area is one. The probabilities computed correspond to percentages of the data.

Example 5. *Translate between probability and percentile questions about scores of students taking an exam.*

2. Normal distributions

Example 6. *[Main Example]* $\mathbf{Z}$ *is a number chosen with normal distribution* $N(\mu, \sigma)$. *We'll use* $N(0,1)$ *as our first example.*

- *Our random event is the choice of a number.*
- *Our random variable* $\mathbf{Z}$ *is the value of the number.*
- *Our sample space is all possible real numbers.*
- *Our probability distribution is the function*

$$e^{-t^2/2}/\sqrt{2\pi}.$$

- *What is* $P(\mathbf{Z} > 3)$*? (Hint: use 68-95-99.7 rule.)*
- *What is* $P(0 \leq Z \leq 2)$*?*
- $P(\mathbf{Z} > 100)$*?*

Doing calculations with normal distributions works much as before. Instead of calculating percentiles, we are calculating probabilities.

Example 7. *Let our random event be the random choice of a U.S. adult male, assuming a probability distribution which is approximately normal* $N(70,4)$*..*

(1) *What is our random variable? Our state space? (theoretical vs. actual)*
(2) *What is the probability that our randomly chosen man is between 70 inches and 73 inches tall?*

3. Sampling distributions

A basic practice in statistics is to take a distribution and *sample* from it.

Definition 8. *Given a population, the sampling distribution for samples of size n is the probability distribution of some parameter as we take values n times.*

We go thoroughly through an example looking at sampling distributions and comment thoroughly as we go along.

Consider the collection of numbers

$$\mathbf{S} = \{2, 2, 5, 6, 7\}.$$

What is the mean? Ans: 4.4.

We can consider all samples from our collection of size 1. There are 5 of them and again, the mean of the samples is again 4.4. The standard deviation is 2.302.

We can now consider all samples from $\mathbf{S}$ of size 2, and take the mean of each sample. Our samples are

$$\{\{2,2\}, \{2,5\}, \{2,6\}, \{2,7\}, \{2,5\}, \{2,6\},$$
$$\{2,7\}, \{5,6\}, \{5,7\}, \{6,7\}\}.$$

Our means are

$$\{2, 3.5, 4, 4.5, 3.5, 4, 4.5, 5.5, 6, 6.5\}$$

Notice that the spread is smaller, even though there are more numbers. The mean is still 4.4, and the standard deviation is now 1.329.

The probability distribution of these numbers is an example of *sampling distribution.* In this case it is for the mean of a sample of 2 from the set $\mathbf{S}$. There are 7 possible outcomes which *don't* have equal probabilities.

X	2	3.5	4	4.5	5.5	6	6.5
Probability	.1	.2	.2	.2	.1	.1	.1

Now consider all samples of size 3, and their means. There are again 10 such samples (you might recall this from our digression on binomial coefficients), which we won't list. The means for these 10 samples are

$$\{3, 3.333, 3.667, 4.333, 4.667, 4.333, 4.667, 5, 5, 6\}$$

The standard deviation is now only .886.

Our sampling distribution for samples of 3 out of our original group S is

X	3	3.333	3.667	4.333	4.667	5	6
Probability	.1	.1	.1	.2	.2	.2	.1

Example 9. *Calculate the sampling distribution for the mean of samples of two out of the data set* $\{-1, 0, 0, 2, 3, 6\}$.

The process of taking a sampling distribution is not an easy one to understand the first time you see it; go through this example and the book carefully.

There are two important cases of sampling distributions. The first is one we have seen in a different language.

Theorem 10. *The sampling distribution measuring number of heads in n coin flips is called the* binomial distribution*, with values giving by Pascal's triangle. As n gets large, this sampling distribution approaches a normal distribution.*

We won't elaborate on this example - this is just recasting our coin flipping observations in a new language. The next example is a new one for us.

Theorem 11. *The sampling distribution for the mean of n measurements of some data which is known to be following a distribution of $N(\mu, \sigma)$ is itself normal, given by $N(\mu, \sigma/\sqrt{n})$.*

Example 12. *For men's heights, distributed according to $N(70, 4)$, what is the probability that nine men chosen at random have an average height between 70 and 73?*

Compare this with our previous answer as to the probability of finding a single man with such height.