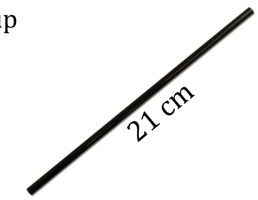
Activity #10: Special continuous distributions

In activity 6, we learned about the probability distributions of discrete random variables. Then, in activities 7-8, we learned about special discrete distributions (binomial, geometric, negative binomial, hypergeometric, and poisson).

In activity 9, we learned about continuous distributions (how to calculate probabilities, find expected values, and percentiles). In this activity, we'll learn about a couple special continuous distributions.

1) Picture a woman waiting to interview for a job. She's nervous, so when she receives a drink in a cup with a straw, she (without thought) bends the straw.

   Let's assume each place along the 21 cm straw has an equal chance of being bent by our nervous interviewee. If we let X = the place along the strong that the interviewee bends, is X a discrete or continuous variable?

   What is the probability that the interviewee bends the straw at a point 13.4 centimeters from the left end?

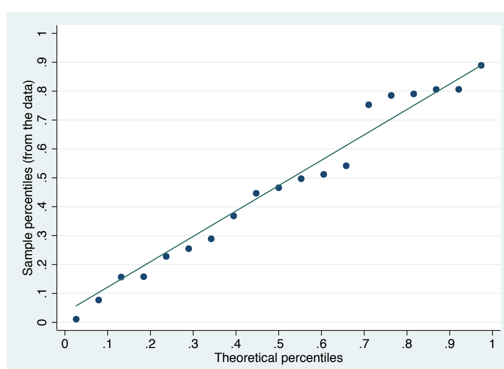
2) How could we estimate the probability that the interviewee bends the straw somewhere between 0 and 10.5 centimeters from the left end? What is the probability that the straw is bent in any 1-cm interval?

In scenarios where the outcome is completely arbitrary, except that we know it lies between certain bounds (from a to b), we have a **uniform distribution.**

pdf: $f(x) = \dfrac{1}{b-a}$ for $a \le x \le b$ cdf: $F(x) = \dfrac{x-a}{b-a}$ $E(X) = \dfrac{a+b}{2}$

3) Here are 19 "random" numbers Microsoft Excel gave me (sorted from smallest to largest). Do these numbers appear to be randomly selected from the interval from 0 to 1? If they were random, then they should represent the entire interval consistently (i.e., the first value should represent the first 5% of the distribution, the second value should represent the 10th percentile, and so on). To check to see if these numbers follow a uniform distribution, I created the following q-q plot. What can you conclude?

Random Values	Theoretical percentile	Random Values	Theoretical percentile
0.010717	0.05	0.496895	0.55
0.076856	0.10	0.511768	0.60
0.156496	0.15	0.541575	0.65
0.157753	0.20	0.752679	0.70
0.227839	0.25	0.784926	0.75
0.254670	0.30	0.790222	0.80
0.288771	0.35	0.805580	0.85
0.367893	0.40	0.805958	0.90
0.446442	0.45	0.889043	0.95
0.465619	0.50		



4) Recall, from activity 8, that we can use the geometric distribution to estimate the probability that the first success happens on the kth trial. In a sense, then, we use the geometric distribution to estimate the probability that we will wait a certain number of trials before something happens.

Instead of dealing with trials, suppose we're interested in the probability that we will wait a certain amount of <u>time</u> before something happens. Let's let X represent the time it takes before an event occurs. To get us started, suppose we know this event happens (on average) λ times in a single unit of time.

Let's divide the single unit of time into *n* very small subintervals of time. How many sub-intervals? Lots. Let's just say *n* is a very large positive number.

With the notation we're developing, the probability that the event occurs during a certain sub-interval of time is λ/n.

Suppose we're interested in P(X ≤ b), which is the probability that the event occurs within the first *b* subintervals of time.

This sounds like a geometric distribution problem, so let's also let Y follow a geometric distribution in which p = P(the event occurs in any trial) = λ/n.

Then, we can say, P(X ≤ *b*) ≃ P(Y ≤ *bn*), which is the probability that it takes *bn* sub-intervals for the event to occur.

We know how to calculate P(Y ≤ *bn*), since it follows a geometric distribution.

$$P(Y \le bn) = \sum_{k=1}^{bn} P(Y = k) = \sum_{k=1}^{bn} (1-p)^{k-1} p = \sum_{k=1}^{bn} \left(1 - \frac{\lambda}{n}\right)^{k-1} \frac{\lambda}{n}$$

Using some algebraic manipulation, we can find:

$$= \sum_{k=1}^{bn} \left(1 - \frac{\lambda}{n}\right)^{k-1} \frac{\lambda}{n} = \frac{\lambda}{n} \sum_{k=1}^{bn} \left(1 - \frac{\lambda}{n}\right)^{k-1} = \frac{\lambda}{n} \sum_{k=0}^{bn-1} \left(1 - \frac{\lambda}{n}\right)^{k}$$

This expression on the right is a finite geometric series. Depending on how much Calculus II you remember, you may recall:

$$\sum_{k=0}^{m} a^k = \frac{1-a^{m+1}}{1-a}$$

In our scenario, we have (from the above formula): *m* = *bn-1* and *a* = *1-λ/n*. That leaves us with:

$$P(Y \le bn) = \frac{\lambda}{n} \left(\frac{1-(1-\lambda/n)^{bn}}{1-(1-\lambda/n)} \right) = \frac{\lambda}{n} \left(\frac{1-(1-\lambda/n)^{bn}}{\lambda/n} \right) = 1 - \left(1 - \frac{\lambda}{n}\right)^{bn} = 1 - \left(\left(1 - \frac{\lambda}{n}\right)^{n} \right)^{b}$$

Again, if you remember your Calculus II class well, you'll recall that as n approaches infinity:

$$P(X \le b) = \lim_{n \to \infty} \left[1 - \left(\left(1 - \frac{\lambda}{n}\right)^{n} \right)^{b} \right] = \lim_{n \to \infty} \left[1 - \left(e^{-\lambda}\right)^{b} \right] = 1 - \left(e^{-\lambda b}\right)$$

What we just derived is the cumulative distribution function, F(X), for the exponential distribution.

5) If you don't like that previous derivation, we can also derive the exponential distribution from the Poisson distribution.

Recall the Poisson distribution models the number of times an event occurs in a given unit of time (or distance, weight, etc.) If we let λ represent the average rate of occurrences per unit of time, then we found:

$$P(X = x) = \frac{\lambda^x}{x!} e^{-\lambda}$$ (This is the probability function for the Poisson distribution)

We'll define an **exponential** random variable, X, as a model for the time we must wait until the next occurrence (or the waiting time).

To calculate the cumulative distribution of X -- P(X ≤ x) -- let x represent a positive value of time:

$$F(x) = P\left(X \le x\right) = 1 - P(X > x) = 1 - P(\text{Time to next occurrence} > x)$$

$$1 - P(\text{Time to next occurrence} > x) = 1 - P(\text{No occurrences in the interval from 0 to } x)$$

$$\begin{aligned} &1 - P(\text{No occurrences in the interval from 0 to } x) \\ &= 1 - P(\text{a Poisson random variable} = 0) \\ &= 1 - \left(\frac{\lambda^0}{0!} e^{-\lambda x}\right) \\ &= 1 - e^{-\lambda x} \end{aligned}$$

When we want to model waiting times (how long we must wait for an event to occur), we can use the **exponential distribution**:

pdf: $f(x) = \lambda e^{-\lambda x}$ cdf: $F(x) = 1 - e^{-\lambda x}$ $E(x) = \frac{1}{\lambda}$

6) Use the pdf to verify the cdf.

7) Suppose you have conducted a study to measure the life expectancy of “Brand X” batteries. Is the variable X = the life expectancy (measured in hundreds of hours) a discrete or continuous variable? What is the probability that a randomly selected battery has a life expectancy of exactly 1 year?

8) Suppose you find the life expectancy of the batteries are modeled by an exponential distribution with $\lambda = 0.44$. Write out and graph the pdf for X. Then, write out and graph the cdf for X.

9) Calculate the following probabilities:

$P(X \leq 3) =$

$P(2 \leq X \leq 8) =$

$P(X \geq 3) =$

10) On average, a bus arrives at a terminal every 10 minutes. What is the probability that a person must wait for more than 20 minutes for the next bus?

11) Suppose Mary arrives at the terminal and finds out John has been waiting for a bus for 10 minutes. What’s the probability that Mary will have to wait more than 20 minutes for the next bus?

12) According to the National Transportation Safety Board, there were 14 plane crashes (not all fatal) in 2007. Using this number, we can assume a plane crashes, on average, every 26 days.

Given we have a plane crash on July 1, what's the probability that we will have another crash in July?

What's the median number of days we must wait until the next plane crash?

What is the variance between accidents? Take the square root of this value and interpret.

For extra credit, choose one of the following distributions:

Multinomial Pareto Gamma Beta Weibull

and do the following:

1) Write out the pdf and cdf
2) Briefly explain when you would use this distribution
3) Provide and solve 2 examples of probabilities you can calculate with the distribution