## Modeling Uncertainty in the Earth Sciences

**review of statistics and probability theory**

**Question 1**

1. What is an inter-quartile range?
2. Why are mean and median the same for a symmetric distribution? (no math proof, just explain)
3. What is the difference between an “extreme” and an “outlier”?

**Question 2**

1. When X is a continuous random variable that is uniformly distributed in the interval [0,1] then
   1. Pr(X=0.5) = ?
   2. Prob(X>0.5) = ?
2. Draw the pdf and cdf of a random variable X that is uniformly distributed in the interval [a,b].
3. Use the cumulative distribution function below to show how you can calculate 1) the median, 2) the 10%-quantile (lowest decile)



1. On the same graph, draw a random outcome of this distribution model using pseudo-random number 0.6.

**Question 3**

A fragment of an unknown species of mosasaur has been found in a river, see figure. A vertebrate paleontologist would like to send a search party to get more complete remains. Unfortunately, the source of the fragment cannot be identified with certainty, because the fossil was found below the junction of two stream tributaries that drain different basins. The drainage basin of the larger stream contains about 18 square miles, while the smaller contains 10 square miles.



The following information is available: 35% of the outcropping of Cretaceous rocks in area 1 are marine, 80% of the outcropping of Cretaceous rocks in area 2 are marine.

1. What is the probability of finding a fossil fragment in each area without knowing this information?
2. What is the probability of finding a fossil fragment in each area with the additional information?

**Question 4**

From nine stations in the Pacific Ocean, you obtained a sample of temperatures in degrees F as follows

x1= 54, x2=65, x3= 66, x4=66, x5=68, x6=70, x7=72, x8=74, x9=81

1. Draw the empirical distribution function.
2. Explain the statement: “Any series of numbers represent a distribution function”
3. Demonstrate graphically how you would transform the histogram of this data into a new distribution *FY(y)* whose mathematical representation is given as follows



1. How would you check if the distribution of this data is uniform in the interval [50, 83]?