1. Consider a set of trees with tags on the nodes. The trees may be parse trees of computer programs, phylogenic trees, or some other type. The trees have thousands of nodes. You would like to define how similar trees are and be able to construct a small sketch of each tree that would allow you to answer the question as to how similar a tree is to any other tree in the collection. How would you do this?
Hint: Use a shingle of some type.
2. Develop a formula for the surface area of a sphere in dimensions.
3. Generate a 1000 pairs of random points on the surface of a hyper sphere in 900 dimensions. Make sure that the points are uniformly distributed on the sphere. What is the distribution of distances between the two points in each pair? Plot the distribution of the distances.
4. (a) Write a computer program that generates $n$ points uniformly distributed over the surface of a d dimensional sphere.
(b) Create a random line through the origin and project the points onto the line. Plot the distribution of points on the line.
(c) What does your result from part b say about the surface area of the sphere in relation to the line, i.e., where is the surface area concentrated relative to the line?
