Homework 7 - Due Mon. Apr. 7th
AS.110.446, EN.550.416

Dr. Mauro Maggioni
Web page: www.math.jhu.edu/˜ mauro
Office: 405 Krieger Hall
E-mail: myfirstname.maggioni@jhu.edu

Homework Policies

As in the first homework set.

Exercises

1. Consider the set X_n^D consisting of n points $x_1, \ldots, x_n$ sampled independently uniformly from the cube $Q^D := [-1,1]^D := \{x \in \mathbb{R}^D : x_i \in [-1,1] \text{ for all } i = 1, \ldots, D\}$. Consider the random variable Y^D given by the Euclidean length of vectors sampled as above. We want to understand the behavior of this random variable, especially for large D, and how its empirical counterpart Y_n^D, obtained by looking at the length of n samples, compares to Y^D. So for each of the choices $D = 8, 16, 32, 64, 128$ and $n = D, 2D, 4D, \ldots, 2^{10}D$, draw $x_1, \ldots, x_n$ as above, consider $y_1 = ||x_1||, \ldots, y_n = ||y_n||$ and consider a histogram approximating the distribution of Y^D. You may use the Matlab histogram function `hist` (or `histc`) to look at a binning of the distribution of the y_i's. As a function of n and D, and especially for D large, how does this distribution behave? What is its (estimated) mean? What is its (estimated) variance? How do these depend on D? Does these estimate seem to depend a lot on n? Why is our smallest choice of n equal to the dimension D? Which proportion of the points is within (Euclidean) distance t from one of the vertices of the cube, as a function of t? Where is the vast majority of points? [25 pts]

2. The same as the above, but for X distributed as a standard normal distribution $\mathcal{N}(0, I_D)$, for $D = 1, 2, 4, 8, 16, 32$ and $n = D, 2D, 4D, \ldots, 2^{10}D$. Y again is the random variable equal to the length of X. Instead of asking about distance from the vertices of the cube, ask yourself what proportion of points has length squared within t of the value D. [25 pts]

3. We would like to do the same as the above, but for the uniform normalized probability measure in the unit ball $\mathbb{B}^D = \{x \in \mathbb{R}^D : ||x|| \leq 1\}$. However sampling efficiently from this distribution is not trivial. In order to sample uniformly from the ball, we shall use an accept/reject method: we sample uniformly in the cube $[-1,1]^D$ (something we can do easily), then check is such a sample is in $\mathbb{B}^D$: if it is we accept it, if it is not we throw it away and try again. Prove/argue that the samples kept are indeed uniformly sampled in $\mathbb{B}^D$. Code this up and try for $D = 2, 4, 8, 16$, and $n = 1000$. When you perform this sampling, keep track of how many samples you keep (let's call this $n = n_{accept}$ and how many you discarded (let's call this n_{reject}). How does $\frac{n_{accept}}{n_{accept}+n_{reject}}$ behave as a function of D? Relate this to a relationship between the volume of $\mathbb{B}^D$ and volume of Q^D. Conclude that this method is extremely inefficient as D grows. [50 pts]