Datasta Tietoon, Autumn 2007

EXERCISE PROBLEMS 1 [Nov 9th 2007, Nov 14th 2007]

H1 / 1. (Convolution filter)
Convolution filtering is computed by the formula

$$g_k = \sum_{m=-\infty}^{\infty} f_m h_{k-m}$$

where f_m is the (discrete) input signal, h_n is the filter sequence, and g_k is the output signal. Compute and plot the output signal when
a)

$$f_0 = 1,\ f_m = 0 \text{ otherwise;} \quad (1)$$
$$h_0 = 2,\ h_1 = 1,\ h_n = 0 \text{ otherwise} \quad (2)$$

b)

$$f_0 = 2,\ f_1 = -1,\ f_m = 0 \text{ otherwise;} \quad (3)$$
$$h_0 = -1,\ h_1 = 2,\ h_2 = 1,\ h_n = 0 \text{ otherwise.} \quad (4)$$

H1 / 2. (Fourier transform and filtering)
In the frequency domain, the convolution formula of Exercise 1 reads

$$G(\omega) = H(\omega)F(\omega)$$

where the functions are Fourier transforms of the corresponding discrete sequences, e.g.,

$$F(\omega) = \sum_{m=-\infty}^{\infty} f_m e^{-i\omega m}.$$

a) Show that the inverse transform is

$$f_n = \frac{1}{2\pi}\int_{-\pi}^{\pi} F(\omega)e^{i\omega n} d\omega.$$

b) The Fourier transform of an ideal low-pass filter (on interval $-\pi \leq \omega \leq \pi$) is

$$H(\omega) = 1 \text{ if } |\omega| \leq \omega_0,\ 0 \text{ otherwise.} \quad (5)$$

By the inverse transform, compute the corresponding sequence h_n and plot it when $\omega_0 = \pi/2$.

H1 / 3. (Substring histograms)
The DNA molecule can be written as a string with four possible letters A, C, G, T, e.g. ...AAGTACCGTGACGGAT ... Assume that the length is a million letters. We want to form histograms for substrings of length n (if $n = 1$, for A, C, G, T; if $n = 2$, for AA, AC, ... TT and so on). How large should n be chosen if we need on the average at least 10 substrings in each histogram bin?

H1 / 4. (High-dimensional spaces)
The d dimensional data vectors are uniformly distributed over a hypercube with side length 1. Inner points are those whose distance from the surface of the hypercube is at least $\epsilon > 0$. Show that the relative volume of the inner points tends to 0 as $d \to \infty$, that is, for very large dimensions almost all the points are on the surface.

H1 / 5. (High-dimensional spaces)
In the lectures, it was stated without proof that the average distance of n points in a d dimensional hypercube is

$$D(d, n) = \frac{1}{2}(\frac{1}{n})^{\frac{1}{d}}.$$

This is an approximate formula. Let us look at a special case: the n points are in the centers of n similar non-overlapping smaller cubes whose union makes up the whole hypercube. Show that then the distances of the points are

$$D(d, n) = (\frac{1}{n})^{\frac{1}{d}}.$$

The distance between two points $\mathbf{x}_1, \mathbf{x}_2$ is defined as $\max_i |x_{i1} - x_{i2}|$. Consider the case $d = 2$, $n = 4$ and verify your result.