MATH475b. Homework 4 — Due May 2nd

In this homework we will study the basics of (pseudo) random number generation.

Linear congruential generators. Recall, that these are defined by recurrence relations,

$$X_{n+1} = (\alpha X_n + \beta) \mod M,$$

for some α (multiplier), β (increment), and M (modulus). As we discussed in class, provided M and β are relatively prime, $\alpha - 1$ is divisible by all prime factors of M and by 4 if M is divisible by 4, the generator will have the full period (equal to M). Take $M = 2^{16}$ and pick some α and β which satisfy the aforementioned conditions on your own.

Uniformly distributed random variables. Your RNG produces integers in the $[0, M-1]$ range. Transform them into $\mathcal{U}[-1,1]$, i.e., uniformly distributed in the interval $[-1,1]$ random numbers by by an appropriate linear transformation.

Q: Plot a bin-counting histogram illustrating your random number generation: split the interval $[-1,1]$ into, e.g., a hundred "bins" — sub-intervals of 0.02 size. Generate the full sequence of $\mathcal{U}[-1,1]$ random numbers (all M of them) and plot a histogram which shows the fraction of numbers which "fell" into each bin. If your RNG has period M, you should get a uniform distribution.

Q: Plot the *auto-correlation function* of your random sequence. It is defined as

$$f(k) = \mathbb{E}(X_n X_{n+k}) = \frac{1}{M}\sum_{n=0}^{M-1} X_n X_{n+k} \quad (n+k \quad \text{should be computed modulo } M).$$

Ultimately, if all the random variables that you generated were truly independent, $f(k)$ would be equal to $2/3$ for $k = 0$ (expected value of the square of a $\mathcal{U}[-1,1]$ random variable), and 0 for all other k. What do you get? *Computing the autocorrelation function may be quite time consuming if you do it directly. However, this is one of the standard application of the DFT: compute the DFT of the entire vector of X_n-s, square it (by absolute value) and compute the inverse DFT. You get the correlation function! Why?*

Gaussian random numbers. Recall, that a *normal* random variable, X, is characterized by the following probability distribution:

$$\mathbb{P}\{X \le x\} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} \mathrm{e}^{-y^2/2}\,\mathrm{d}y =: F(x). \qquad \text{We say, } X \text{ is } \mathcal{N}(0,1).$$

In order to generate normal random variables, first generate uniform random points in a unit disk centered at the origin. Do this either using "draw and discard" method by generating points in the $[-1,1]^2$ box and discarding those outside of the unit disk, or using polar coordinates directly: $r = \sqrt{\mathcal{U}(0,1]}$, $\theta = \mathcal{U}[0,2\pi]$. Then $2\sqrt{-\ln r}\cos\theta$ and $2\sqrt{-\ln r}\sin\theta$ are normal (independent) random variables.

Q: Generate a few thousands $\mathcal{N}(0,1)$ random points and plot a bin-counting histogram, e.g., use 100 bins over the interval $[-5,5]$, to verify your algorithm. Compare your histogram to the function $\mathrm{e}^{-x^2/2}/\sqrt{2\pi}$. *(You may use either your own or built-in RNG for this exercise.)*