DATA, MODELS & UNCERTAINTY IN THE NATURAL SCIENCES

Problem Set 2

In this lab we will work with *probability density functions* (`pdf`) and *probability distribution functions* or *cumulative distribution functions* (`cdf`) and their inverses (`icdf`). Many of them are preprogrammed in MATLAB—probably all you may ever need (assuming you have the `stats` toolbox installed, which comes with the Princeton distribution).

You can convince yourself that this all works by trying `icdf('normal',0.975,0,1)` and observing that you get... you guessed it: 1.96. Indeed, 95% of a normal distribution with a mean $\mu = 0$ is contained within $\pm 1.96\sigma$. For the same reason, `cdf('normal',1.96,0,1)` evaluates to 0.975. I hope you understand why.

We will "verify" the **Central Limit Theorem** in its most general form. Let's verify for ourselves, heuristically, that indeed, to speak with *Bendat & Piersol* (3^{rd} edition, pp 65–67):

> The normal distribution will result quite generally from the sum of a large number of independent random variables acting together. Let $x_1(k), x_2(k), \ldots, x_N(k)$ be N mutually independent random variables *whose individual distributions are not specified and may be different*. Let μ_i and σ_i^2 be the mean value and variance of each random variable $x_i(k)$, $i = 1, 2, \ldots, N$. Consider the sum random variable
>
> $$x(k) = \sum_{i=1}^{N} a_i x_i(k), \tag{1}$$
>
> with a_i arbitrary fixed constants. The mean value μ_x and the variance σ_x^2 become
>
> $$\mu_x = \sum_{i=1}^{N} a_i \mu_i, \tag{2}$$
>
> $$\sigma_x^2 = \sum_{i=1}^{N} a_i^2 \sigma_i^2. \tag{3}$$
>
> The Central Limit Theorem states that under fairly common conditions the sum random variable $x(k)$ will be normally distributed as $N \to \infty$ with the above mean value μ_x and variance σ_x^2.

As we know, the probability density function $p_Z(z)$ of $z = x + y$, the sum of two independent random variables is given by the convolution of their respective probability density functions (*Bendat & Piersol*, p. 61):

$$p_Z(z) = p_X * p_Y = \int_{-\infty}^{+\infty} p_X(x)\, p_Y(z-x)\, dx. \tag{4}$$

1. Illustrate computationally that successive convolutions of uniform distributions tend to Gaussian behavior in the manner described above. You may use `conv` and `normpdf`.
2. Illustrate that successive convolutions of almost any function that is a proper probability density function tend to Gaussian behavior. You may want to use the functions `trapz` or `quad` to carry out the numerical integrations to find the moments $\int p(x)\, dx$, $\int xp(x)\, dx$ and $\int x^2 p(x)\, dx$ of the distributions you specify.