M1 - Marine Physics **2019–2020**

Data Analysis

#2 Central Limit Theorem

1 Mean as a random variable

1.1 Normal distribution

We define a n-sample as `n` realizations of a normally distributed random variable.

1. Write a `for` loop that you iterate `nsamp` times in which you define a n-sample and compute its mean $\bar{x}$. Use $n = 500$.
2. Use `nsamp=4000` and plot the PDF of $\bar{x}$.
3. $\bar{x}$ is itself a random variable. What is its mean ? its standard deviation $\sigma_{\bar{x}}$? Does the distribution look Gaussian ?
4. Check how the pdf and standard deviation $\sigma_{\bar{x}}$ vary with n.

1.2 chi-squared distribution

We now redo exactly the same computation using a random variable following a chi-squared distribution

1. Write a function to generate a random variable following a chi-squared distribution with $l = 10$ degrees of freedom
2. Write a `for` loop that you iterate `nsamp` times in which you define a n-sample (following a chi-squared distribution) and compute its mean $\bar{x}$. Use $n = 500$.
3. Use `nsamp=4000` and plot the PDF of $\bar{x}$. Does the distribution of the means still look Gaussian ?
4. Check how the pdf and standard deviation $\sigma_{\bar{x}}$ vary with n.

1.3 Cauchy distribution

We now redo exactly the same computation using a random variable following a Cauchy distribution

1. Write a function to generate a random variable following a Cauchy distribution
2. For different n, compute the estimated standard deviation and show that it's diverging with n, the number of samples.
3. Write a `for` loop that you iterate `nsamp` times in which you define a n-sample (following a Cauchy distribution) and compute its mean $\bar{x}$. Use $n = 500$.
4. Use `nsamp=4000` and plot the PDF of $\bar{x}$. Does the distribution of the means still look Gaussian ?
5. Check how the pdf and standard deviation $\sigma_{\bar{x}}$ vary with n.

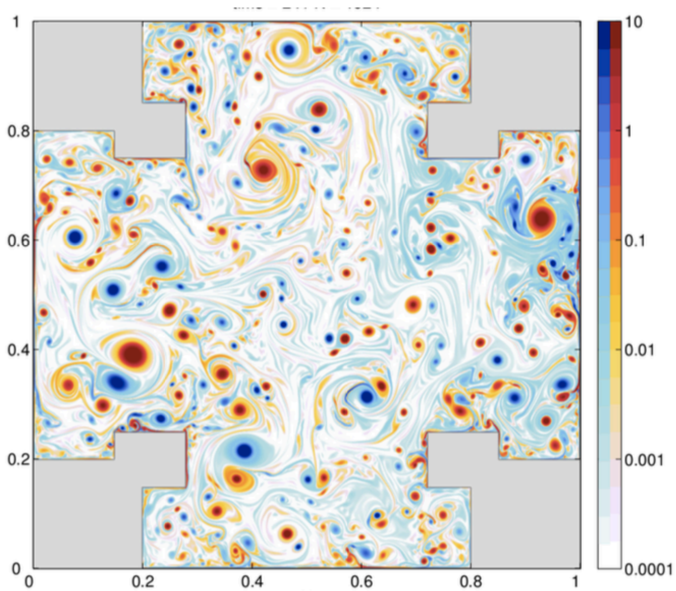
2 Standard deviation as a random variable

We define a n-sample as `n` realizations of a normally distributed random variable as in the first part, and redo the same computation but for σ the estimated standard deviation of x.

1. generate `nsamp=4000` values of the of random variables `x=randn(n,m);`

2. compute the m standard deviations : `sigma=std(x,0,1);`
3. plot the pdf of `sigma`, its cdf and print its first four moments.
4. Check how the pdf varies with n.

3 Intermittency



Not all random variables are gaussian. Here is an example of a very intermittent variable. Load the `turbulence2D_with_boundaries.mat` file. It contains 4 timeseries K (the total kinetic energy), V (the total enstrophy) and their time derivatives with

$$K = \frac{1}{2}\int_{\mathcal{D}} (u^2 + v^2)\, dA \ , \tag{1}$$

$$V = \frac{1}{2}\int_{\mathcal{D}} \zeta^2 \, dA \tag{2}$$

and ζ the vorticity. These timeseries come from a numerical simulation of decaying 2D turbulence in a closed domain $\mathcal{D}$ with friction on the boundaries.

1. plot K and dK/dt as a function of time [use distinct plots]. Are the series stationnary ?
2. same for V and dV/dt.
3. We will focus on the time derivatives. Define `x` as `dVdt` over the time interval [2.10^4 3.10^4]. [use `find`]
4. Plot the histogram of x. Redo the same plot in semilogy axis.
5. Compute the mean, standard deviation, skewness and the kurtosis of x.
6. Superimpose the gaussian that has the same mean and the same standard deviation. Compare the tails (extreme fluctuations).