<u>How to infer the statistics of the solution for an inverse Problem?</u>

We use the linear regression problem, i.e: given a set of points {Xi}, {Yi}
Find the coefficient (a,b) of the straight line that fits (represents) the data
Y = a*x + b

1) **<u>First solution</u>**, run MCMC on the inverse problem and study the distribution of the solutions {a,b} around the maximum likelihood.

   Run the linear_regression_MCMC script after having defined your own logprior and loglikelyhood probability functions (my_problemLR script)
   - For the different kind of probability functions that you may have tried (gaussian, exponential, …) give the <u>histograms</u>, <u>means</u>, <u>variance</u> of the solution, and your estimation of the posterior density functions shape for ***a*** and ***b*** (is it Gaussian, exponential, lognormal, poisson, …….)

2) **<u>Second solution</u>**, when MCMC is not appropriate to solve the problem, one can try the **<u>bootstrap</u>** or **<u>jackknife</u>** **<u>methods</u>**: create a set of solutions by changing a subset of the data. Here we look at the bootstrap method.

   Go back to the course and the least-square solution for the linear regression problem. Use regularization if needed.

   One you have created a dataset of N points {Xi}, {Yi} with random noise, you will remove randomly 10% of the N {Xi} and {Yi} values, and replace them with a duplicate of the 90% others.

   Ex: N=100, we remove the last 10 values from i=91 to i=100, and replace them by the 10 first values i=1 to i=10.
   By doing so, you keep the same number of points: 100, but some points are duplicated.

   Run the least-square inversion, keep the results and retry again with a different subset of data. Do this several times, you will have a set of says 100 solutions from different selection of {Xi}, {Yi} values.
   Study the statistics of these solutions (histogram, mean, variance) and compare to the results of the first solution.