exercise

Let consider n iid random variables $X_1, \ldots, X_n$ whose density is given by

$$f_\theta(x) = (k+1)\theta^{-k-1}x^k 1_{[0,\theta]}(x)$$

Where k is a nonnegative integer fixed and θ is an unknown positive parameter.
In the following, we denote $m = n(k+1)$.

- Determine the the maximum likelihood method, the estimator $\hat{\theta}$.

- Compute the bias of it

- Determine λ_1 and λ_2 such that $\lambda_1.\hat{\theta}$ is unbiased and $\lambda_2.\hat{\theta}$ has the smallest quadratic error.

Linear regression : project

The linear model is one of the most used model in statistics to find a link between a response variable and explanatory ones.
We are looking this model more in details in the sense that I am giving to you a data set and you are trying to answer to some demands.

Our aim is to focus onto the prediction for new observations and to quantify in some sense the accuracy of it.
So, you split the dataset in two parts at random, a first one named training set and a second one named validation set.

Thanks to the training set, estimate the parameters of the linear model, at first by using all the explanatory variables.
Then, let consider the validation set. For each individual of it, compute the associated prediction with the model build previously.
You can compute the test error associated to this validation set.

Make several splittings in two parts to compare this test error. And each time, compare it to the training error.
Which error is better and why?
Now, give the confidence interval for the observation of the response variable associated to a new individual for who we now the values of the explanatory variables.
What are the assumptions needed to compute this confidence interval and how to see if they are satisfied in practice?

Now, try to identify the explanatory variables that are really needed to obtain correct prediction.