Analytics for Observational Data (IT142IU)

Lab 7: Bayesian statistics

## Objectives

* Understanding Bayes’ theorem, Bayesian inference
* Applying Bayesian inference to the existing datasets.
* Dataset sources:
  + <https://www.kaggle.com/datasets/fedesoriano/wind-speed-prediction-dataset>
  + <https://www.kaggle.com/berkeleyearth/climate-change-earth-surface-temperature-data>
* Programming languages: Python/Java
* Ref: Lecture notes in Session 10

## Tasks

|  |  |
| --- | --- |
| **Questions** | **Answers** |
| Dataset | Wind\_dataset.csv |
| Reuse the random variable chosen in the previous lab. | Choosing WIND |
| Choose a good sample from the previous lab |  |
| Calculate Mean, variance, and number of the items in the sample data |  |
| Take TWO items and give their prior distributions for the mean value.  E.g. *p*1(*μ*) ~ *N*(11, 25) for the Wind |  |
| Construct the posterior of the two cases above |  |
| Visualize the distributions of the two cases above |  |
| Remark |  |

|  |  |
| --- | --- |
| **Questions** | **Answers** |
| Dataset | GlobalLandTemperaturesByCountry.csv |
| Choose a random variable | AverageTemperature |
| Choose a good sample from the previous lab |  |
| Calculate Mean, variance, and number of the items in the sample data |  |
| Take TWO items and give their prior distributions for the mean value.  E.g. *p*1(*μ*) ~ *N*(18, 100) for the average temperature |  |
| Construct the posterior of the two cases above |  |
| Visualize the distributions of the two cases above |  |
| Remark |  |