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/carrie1/ecommerce-data>
  + <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 | ecommerce-data |
| Reuse the random variable chosen in the previous lab. |  |
| Choose a good sample from the previous lab |  |
| Calculate Mean, variance, and number of the items in the sample data | 98.88  106269.15  *n* = 1000 |
| Take TWO items and give their prior distributions for the mean value.  E.g. *p*1(*μ*) ~ *N*(500, 2000) for the Total price | Prior 1:    Prior 2: |
| 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 | <give the partial of data here> |
| Calculate Mean, variance, and number of the items in the sample data | 17.2  119.65  *n* = 1000 |
| Take TWO items and give their prior distributions for the mean value.  E.g. *p*1(*μ*) ~ *N*(23.5, 9) for the average temperature |  |
| Construct the posterior of the two cases above |  |
|  |
| Visualize the distributions of the two cases above |
|  |  |
| Remark |  |