

Report on learning practice # 3

Sampling of multivariate random variables

Performed by
Alexander Razin
Mikhail Lovtskiy
Mark Evgrafov
Julia Pimkina
Ac. group J4132c

Table of contents:

1. Substantiation of chosen sampling.

```
df = pd.read_csv(path_to_file)
df = df[ [
    'systolic_blood_pressure',
    'diastolic_blood_pressure',
    'heart_rate',
    'highest_heart_rate',
    'lowest_heart_rate',
    'lowest_urine_output',
    'highest_ph',
    'hematocrit',
    'platelet_count',
    'potassium',
    'lowest_mean_arterial_pressure',
    'highest_mean_arterial_pressure',
    'temperature'
]]
```

Fig.1. Dataset preparation

In this lab, 13 features will be considered, 3 of those are target variables and 10 are predictors (including 1 classifier).

From Lab 2: “Our dataset is composed of a curated collection of over 200 publicly available COVID-19 related datasets from sources like Johns Hopkins, the WHO, the World Bank, the New York Times, and many others. It includes data on a wide variety of potentially powerful statistics and indicators, like local and national infection rates, global social distancing policies, geospatial data on movement of people, and more.”

2. Sampling of chosen target variables using univariate parametric distributions (from practice #2) with 2 different sampling methods.

Based on the classes, we chose Inverse Transform and Accept-Reject methods to sample our target variables. The results are as follows:

1) Base data distribution:

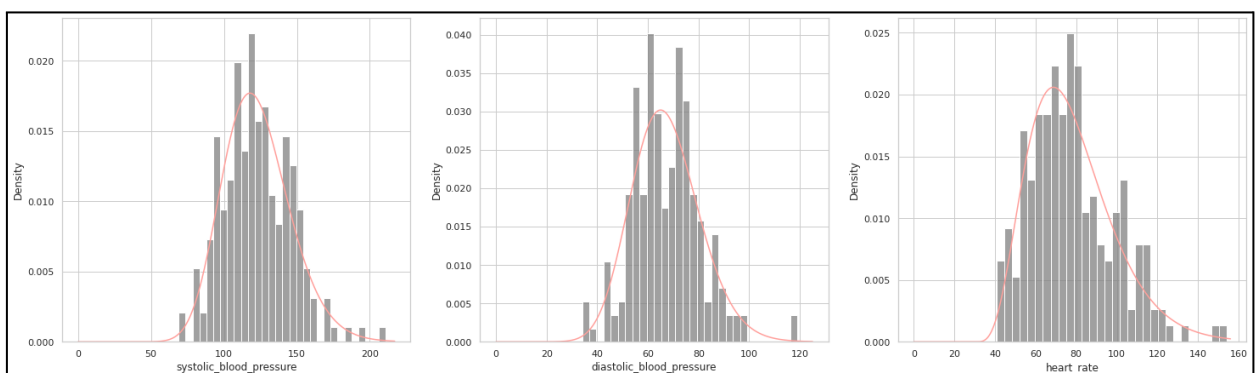


Fig.2. Base data distribution

2) Inverse Transform sampling:

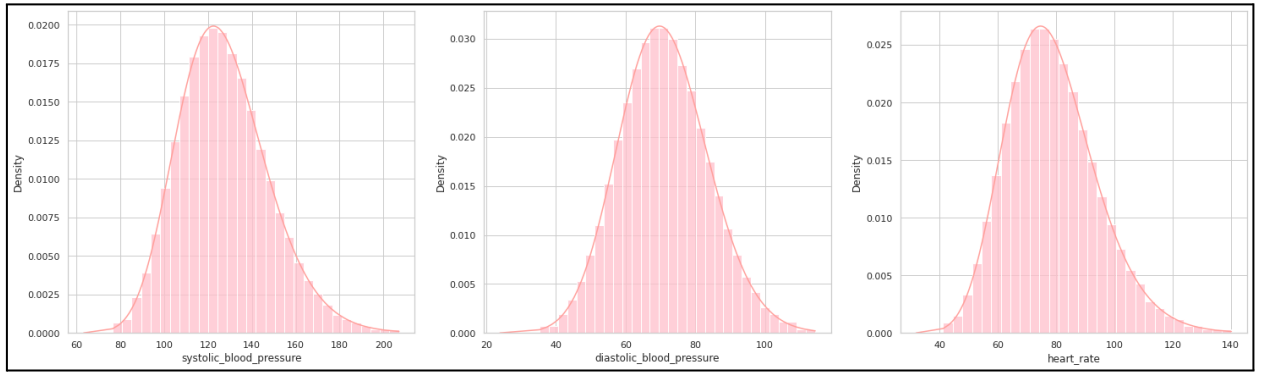


Fig.3. Inverse Transform sampling

3) Accept-Reject sampling:

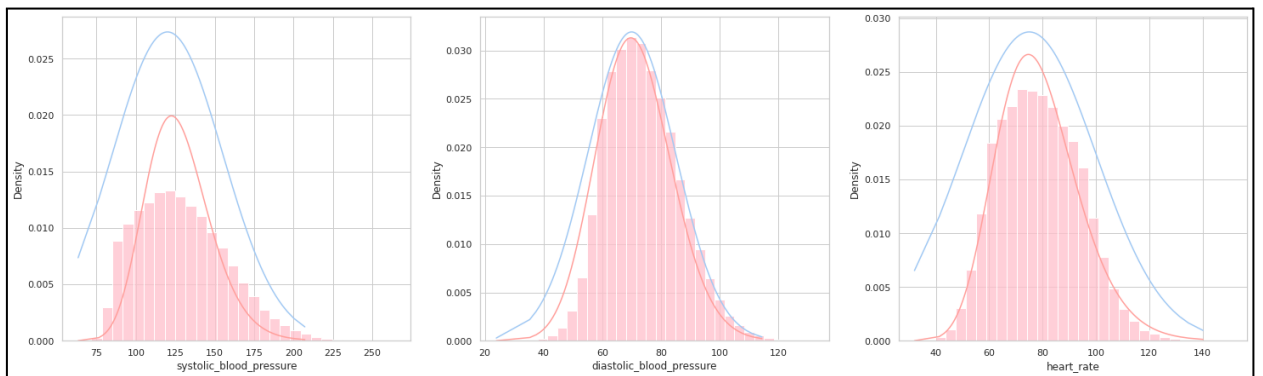


Fig.4. Accept-Reject sampling

3. Estimation of relations between predictors and chosen target variables.

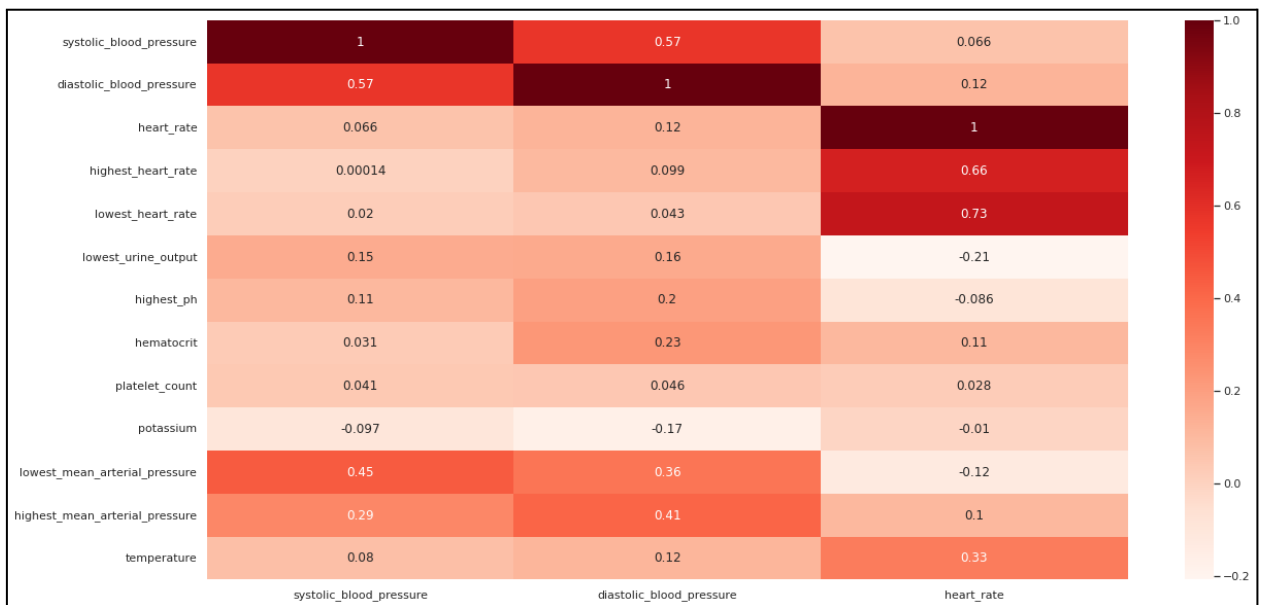


Fig.5. Correlation between predictors and target variables

4. Build a Bayesian network for a chosen set of variables. Choose its structure on the basis of multivariate analysis and train distributions in nodes using the chosen algorithm.

For the Bayesian networks, we used the *pgmpy* library. Based on the correlation matrix, relations between variables for the Bayesian network were built:

```
connections = {
    'temperature': [],
    'diastolic_blood_pressure': ['temperature'],
    'systolic_blood_pressure': ['temperature', 'diastolic_blood_pressure'],
    'highest_heart_rate': ['systolic_blood_pressure', 'temperature'],
    'lowest_heart_rate': ['highest_heart_rate', 'systolic_blood_pressure', 'diastolic_blood_pressure', 'temperature'],
    'heart_rate': ['temperature', 'highest_heart_rate', 'systolic_blood_pressure', 'diastolic_blood_pressure', 'lowest_heart_rate'],
}

edges = create_edges(connections)
vertices = ['temperature', 'diastolic_blood_pressure', 'systolic_blood_pressure', 'highest_heart_rate', 'lowest_heart_rate']
bn = {"V":vertices, "E":edges}
```

Fig.6. Manually created edges

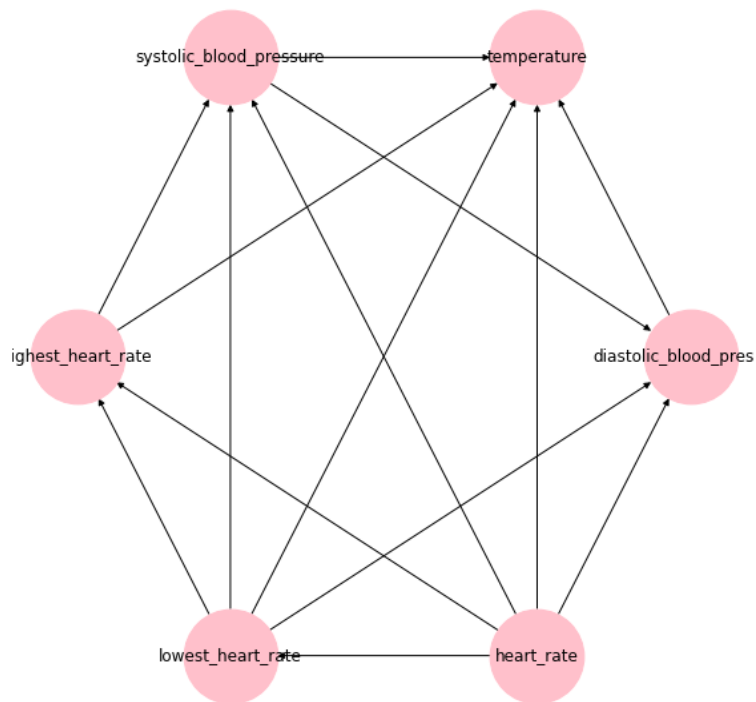


Fig.7. Structure of manually created Bayesian network

5. Build a Bayesian network for the same set of variables but using 2 chosen algorithms for structural learning

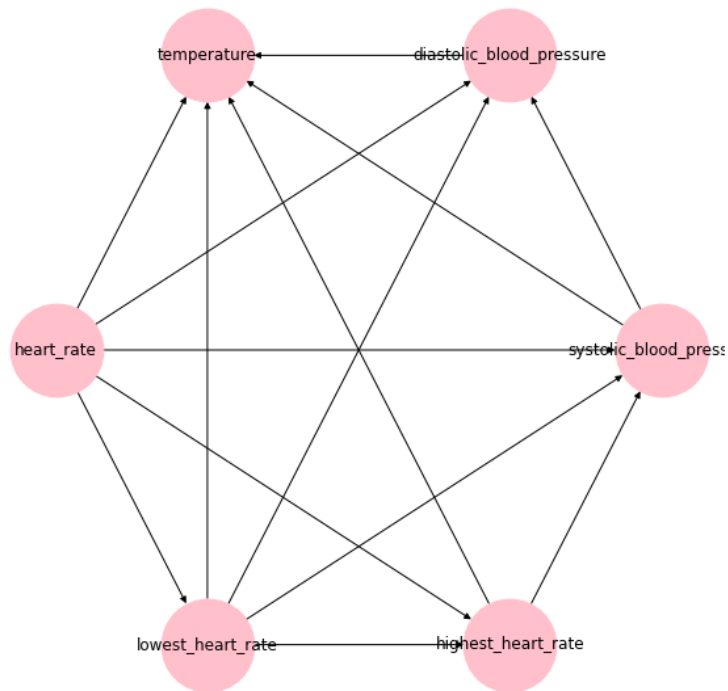


Fig.8. Bayesian network structure with Hill Climb Search using K2Score

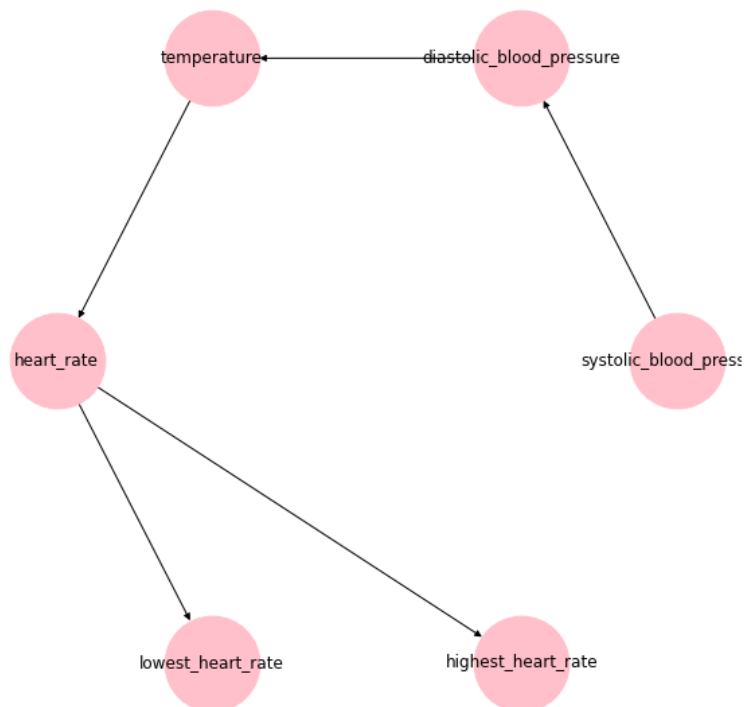


Fig.9. Bayesian network structure with Hill Climb Search using BicScore



Fig.10. Bayesian network structure with TreeSearch

6. Quality analysis.

Due to the task results of *K2Score* and *BicScore* were compared:

1) For *lowest_heart_rate*:

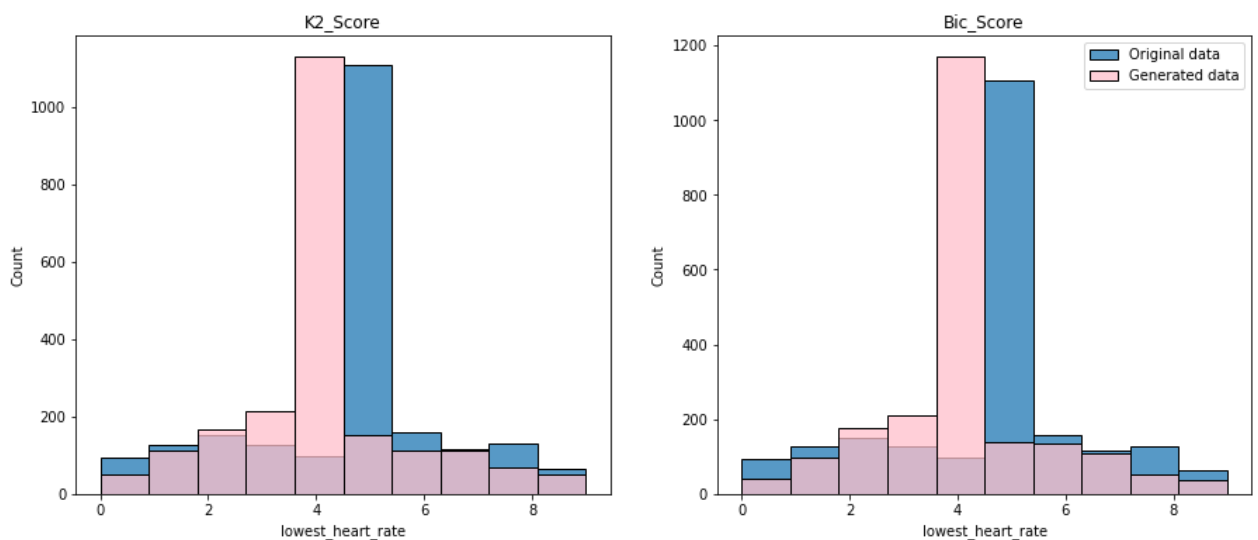


Fig.11. Comparison for *lowest_heart_rate*

2) For *temperature*:

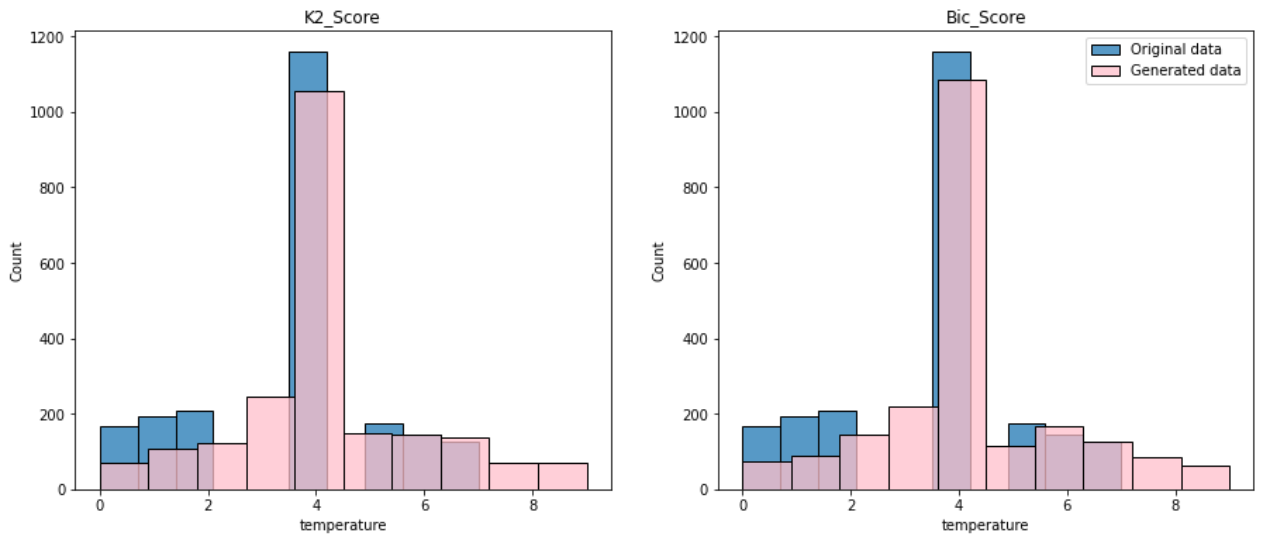


Fig.12. Comparison for temperature

3) For *highest_heart_rate*:

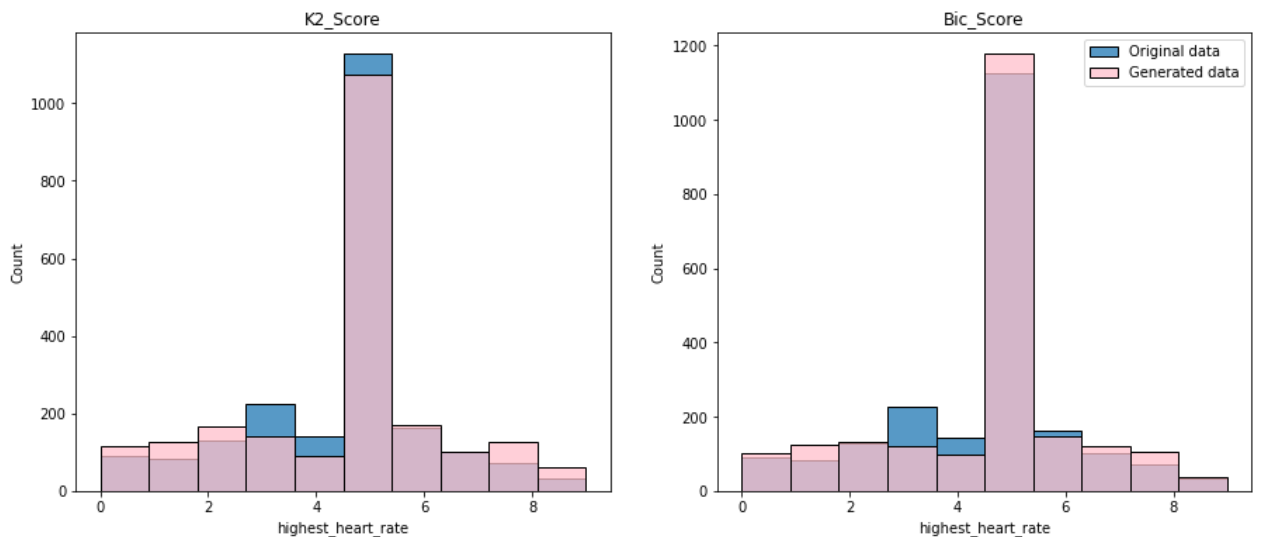


Fig.13. Comparison for highest_heart_rate

Source Code

- The full repository with all the labs: <https://github.com/RazinAleksandr/M-M-MSA-ITMO>
- The repo with Datasets and additional used Data info: <https://github.com/RazinAleksandr/M-M-MSA-ITMO/tree/main/Datasets>
- The Lab3 ipynb files (2 parts): [Part 1](#) (tasks 1-3) & [Part 2](#) (tasks 4-6)

Furthermore, you can find README file with links for every lab folder on the main GitHub repository.