Bayesian Diagnosis

A Bayesian-Based Computational Tool for Parametric and Nonparametric Medical Diagnosis

Version 1.0.0

Theodora Chatzimichail, MRCS [a], Aristides T. Hatjimihail, MD, PhD [b]

Hellenic Complex Systems Laboratory, Kostis Palamas 21, Drama 66131, Greece, [a]tc@hcsl.com, [b]ath@hcsl.com

Interface of the program Diagnostic Uncertainty

About the program controls

The numerical settings are defined by the user with menus or sliders. Sliders can be finely manipulated by holding down the alt key or opt key while dragging the mouse. They be even more finely manipulated by also holding the *shift* and/or *ctrl* keys.

Dragging with the mouse rotates the three-dimensional plots, while dragging with the mouse while pressing the *ctrl*, *alt*, or *opt* keys zooms in or out.

Range of input parameters

v: 0.010 – 0.500

μ: 0.01 – 10000.00

σ : 0.01 – 3000.00

ρ : –1.000 – 1.000

h : 0.01 – 2.00 (standard deviation units)

x: 0.01 – 10000.00

y: 0.01 – 10000.00

Input and output

This program consists of three primary modules with eighteen submodules. It allows the calculation, plotting and comparison of Bayesian posterior probability for disease for two diagnostic tests, assuming two sets of alternative parametric and nonparametric distributions of the measurements of the tests in diseased and nondiseased populations.

Input

For each module the user defines:

1. The prior probability of disease (prevalence)
2. Two sets of four univariate and two bivariate distributions:

    2.1. Two sets of parametric distributions for the parametric module.
    2.2. Two sets of kernel density estimators (KDE) for the KDE module
    2.3. One set of parametric distributions and one set of KDE for the parametric vs KDE module
3. A pair of values (x,y) of the two measurands.

Each set contains the following distributions:

1. The univariate distribution of the first measurand in the diseased population
2. The univariate distribution of the first measurand in the nondiseased population
3. The univariate distribution of the second measurand in the diseased population
4. The univariate distribution of the second measurand in the nondiseased population
5. The bivariate distribution of the first and second measurand in the diseased population
6. The bivariate distribution of the first and second measurand in the nondiseased population

Each univariate parametric distribution is defined by:

1. The type of the distribution:
    1.1. Normal
    1.2. Lognormal
    1.3. Gamma
2. The mean of the measurand in the population
3. The standard deviation of the measurand in the population

Each bivariate parametric distribution is defined by:

1. The two marginal univariate distributions of each measurand in the population
2. The correlation coefficient of the two measurands in the population

For each univariate KDE the user defines its bandwidth (h).

Each bivariate KDE is defined by:

1. The two marginal univariate KDE of each measurand in the population
2. The correlation coefficient of the two measurands in the population

Output

Each module generates:

Plots

The following plots are generated:

Posterior probability for disease:

1. The posterior probability for disease for each measurand and their combination, for the two sets of distributions
2. The difference between the posterior probability for disease for each measurand of and their combination, of the two sets of distributions

Probability density function:

1. The probability density function (PDF) of each measurand and their combination, for the two sets of distributions
2. The PDF of each measurand and their combination, for the two sets of distributions, with the respective histograms of the provided datasets
3. Q-Q plots:

Q-Q plots of each measurand in diseased and nondiseased vs the respective dataset.

4. P-P plots

   P-P plots of each measurand in diseased and nondiseased vs the respective dataset.

Tables

Each module generates:

1. A table with the statistics for each set of distributions and the datasets:
   1.1. Mean
   1.2. Median
   1.3. Standard Deviation
   1.4. Skewness
   1.5. Kurtosis
   1.6. Log-likelihood
   1.7. Correlation coefficient
2. A table with the probability for disease for each set of distributions, and their difference:
   2.1. The prior probability for disease (prevalence)
   2.2. The probability for disease for the first measurand
   2.3. The probability for disease for the second measurand
   2.4. The probability for disease for both measurands, combined

Datasets

The datasets of the program have been obtained from the database of the National Health and Nutrition Examination Survey (NHANES), Centers for Disease Control and Prevention, USA. They are the following:

d1: First measurand (fasting plasma glucose, mg/dl) in diseased (diabetics)

nd1: First measurand (fasting plasma glucose, mg/dl) in nondiseased (nondiabetics)

d2: Second measurand (glycated hemoglobin A1c, %) in diseased (diabetics)

nd2: Second measurand (glycated hemoglobin A1c, %) in nondiseased (nondiabetics)

They are editable.

Notation

Parameters

v: prevalence of disease

μ: mean

σ: standard deviation

ρ_i: correlation coefficient

h: nonparametric kernel density bandwidth

Functions

P(disease): posterior probability for disease

Abbreviations

PDF: probability density function

KDE: kernel density estimator

OGTT: oral glucose tolerance test

FPG: fasting plasma glucose

HbA1c: glycated hemoglobin A1c

NHANES: National Health and Nutrition Examination Survey

About the program

Version

1.0.0

Source Code

Programming language: Wolfram Language
Software source code file format: Wolfram Notebook
Availability: The updated source code is available at:
https://www.hcsl.com/Tools/BayesianDiagnosis/BayesianDiagnosis.nb

Software Requirements

Operating systems: Microsoft Windows, Linux, Apple iOS
Other software requirements: Wolfram Player®, freely available at: https://www.wolfram.com/player/ or Wolfram Mathematica®.

System Requirements

Processor: Intel Core i9® or equivalent CPU
System memory (RAM): 32 GB+ recommended.

Permanent citation

Chatzimichail T., Hatjimihail AT. *Bayesian Diagnosis*: A Bayesian-Based Computational Tool for Parametric and Nonparametric Medical Diagnosis. Drama, Greece: Hellenic Complex Systems Laboratory, 2023. Available at: https://www.hcsl.com/Tools/BayesianDiagnosis/

License: