

Article

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

Theodora Chatzimichail  and Aristides T. Hatjimihail * 

Hellenic Complex Systems Laboratory, Kostis Palamas 21, 66131 Drama, Greece; tc@hcsl.com

* Correspondence: ath@hcsl.com; Tel.: +30-2521058227

Abstract: Medical diagnosis is the basis for treatment and management decisions in healthcare. Conventional methods for medical diagnosis commonly use established clinical criteria and fixed numerical thresholds. The limitations of such an approach may result in a failure to capture the intricate relations between diagnostic tests and the varying prevalence of diseases. To explore this further, we have developed a freely available specialized computational tool that employs Bayesian inference to calculate the posterior probability of disease diagnosis. This novel software comprises of three distinct modules, each designed to allow users to define and compare parametric and nonparametric distributions effectively. The tool is equipped to analyze datasets generated from two separate diagnostic tests, each performed on both diseased and nondiseased populations. We demonstrate the utility of this software by analyzing fasting plasma glucose, and glycated hemoglobin A1c data from the National Health and Nutrition Examination Survey. Our results are validated using the oral glucose tolerance test as a reference standard, and we explore both parametric and nonparametric distribution models for the Bayesian diagnosis of diabetes mellitus.

Keywords: Bayesian diagnosis; Bayesian inference; prior probability; posterior probability; likelihood; parametric distribution; nonparametric distribution; copula distribution; kernel density estimator; probability density function; diabetes mellitus



Citation: Chatzimichail, T.; Hatjimihail, A.T. A Bayesian Inference Based Computational Tool for Parametric and Nonparametric Medical Diagnosis. *Diagnostics* **2023**, *13*, 3135. <https://doi.org/10.3390/diagnostics13193135>

Academic Editor: Roberto Passera

Received: 7 September 2023

Revised: 28 September 2023

Accepted: 29 September 2023

Published: 5 October 2023

Correction Statement: This article has been republished with a minor change. The change does not affect the scientific content of the article and further details are available within the backmatter of the website version of this article.



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Medical diagnosis is a critical process of accurately identifying pathological conditions in patients. The term “diagnosis” has its etymological origins in the ancient Greek word “διάγνωσις”, signifying “discernment” [1]. Traditionally, diagnostic tests are used to divide individuals into two principal categories: those who are afflicted with a specific disease and those who are not. Notably, the probability distributions associated with quantitative diagnostic test outcomes often demonstrate some overlap between the diseased and nondiseased groups. To address this, numerical diagnostic thresholds or cut-off points have been formulated to provide a binary classification of these test outcomes [2]. Nevertheless, this introduces a certain measure of uncertainty into the diagnostic accuracy of those tests [3]. This dichotomous method represents a significant shift in medical decision-making by linking a continuum of evidence to binary clinical decisions such as to treat or not to treat [4].

Despite the evident efficiency of traditional diagnostic methods, they sometimes fail to capture the complexity and heterogeneity of disease presentations across diverse populations [5]. To address these limitations, our research focuses on implementing Bayesian inference to calculate the posterior probabilities associated with disease diagnosis [6–9]. Within this Bayesian paradigm, prior probabilities of disease are integrated with distributions of diagnostic measurands in both diseased and nondiseased populations. This approach enables the evaluation of the information conveyed via diagnostic measurements and the combination of data from multiple diagnostic tests, which may improve diagnostic accuracy and precision while introducing flexibility, adaptability, and versatility

into the diagnostic process [10]. Furthermore, the Bayesian approach extends its utility beyond the medical field by offering a robust framework for quantifying uncertainty in various domains, thereby enriching its applicability in both diagnostic and prognostic contexts [11,12].

A considerable challenge in integrating Bayesian inference into medical diagnosis is the limited availability of literature detailing the statistical distributions of diagnostic variables in both pathological and non-pathological states [13].

The ubiquitous application of the normal distribution in clinical laboratory indicators is due, in part, to its mathematical simplicity, the foundational Central Limit Theorem, and a rich collection of statistical methods designed for Gaussian data [14]. However, the universal applicability of the normal distribution is subject to critique, especially when dealing with clinical measurands that exhibit skewness, bimodality, or multimodality [15]. Hence, while the normal distribution remains invaluable in statistical modeling, critical evaluation of its appropriateness for specific diagnostic measurands is necessary. This evaluation should be accompanied by an openness to adopt alternative statistical distributions when needed [16].

This foundational data is crucial for Bayesian inference, establishing the essential context against which new diagnostic measurements can be compared. The absence of such normative data could potentially compromise the reliability and validity of Bayesian diagnostic methods.

To address the complex issues related to Bayesian diagnosis and the selection of appropriate statistical distributions for diagnostic variables, we have developed the *Bayesian Diagnosis* program, an interactive software tool programmed in the Wolfram Language. This tool allows users to explore and compare both parametric and nonparametric distributions to calculate posterior probabilities for disease. It is designed to analyze datasets of measurements of two distinct diagnostic tests, performed on both diseased and nondiseased populations.

2. Methods

2.1. The Program

Bayesian Diagnosis was developed using Wolfram Mathematica® Ver. 13.3 (Wolfram Research, Inc., Champaign, IL, USA (2023)). This interactive program consists of three primary modules with eighteen submodules. It allows the calculation, plotting, and comparison of Bayesian posterior probabilities of disease for two diagnostic tests, assuming two sets of alternative parametric and nonparametric distributions of the measurements of those tests in diseased and nondiseased populations (refer to Figure 1 and to Supplementary File S1). It is freely available as a Wolfram Notebook (.nb) (Supplementary File: BayesianDiagnosis.nb). It can be run on Wolfram Player® or Wolfram Mathematica® (refer to Appendix B).

Datasets

As datasets are considered tuples of data. Although the program includes four datasets of measurements, one for each diagnostic test, applied to a diseased and a nondiseased population, these can be replaced by other appropriate datasets selected by the user (refer to Appendix B). Therefore, it can be used for any combination of diagnostic tests and diseases.

2.2. Computational Methods

2.2.1. Bayesian Diagnostic Approach

The Bayesian diagnostic approach is a cornerstone in statistical inference and particularly useful in medical diagnosis [6,17,18]. The approach relies on Bayes' theorem [7]. For effective implementation of the Bayesian diagnostic method, knowledge concerning the statistical distributions of the measurements of the diagnostic tests is essential [14].

Bayes theorem is presented in Appendix A.

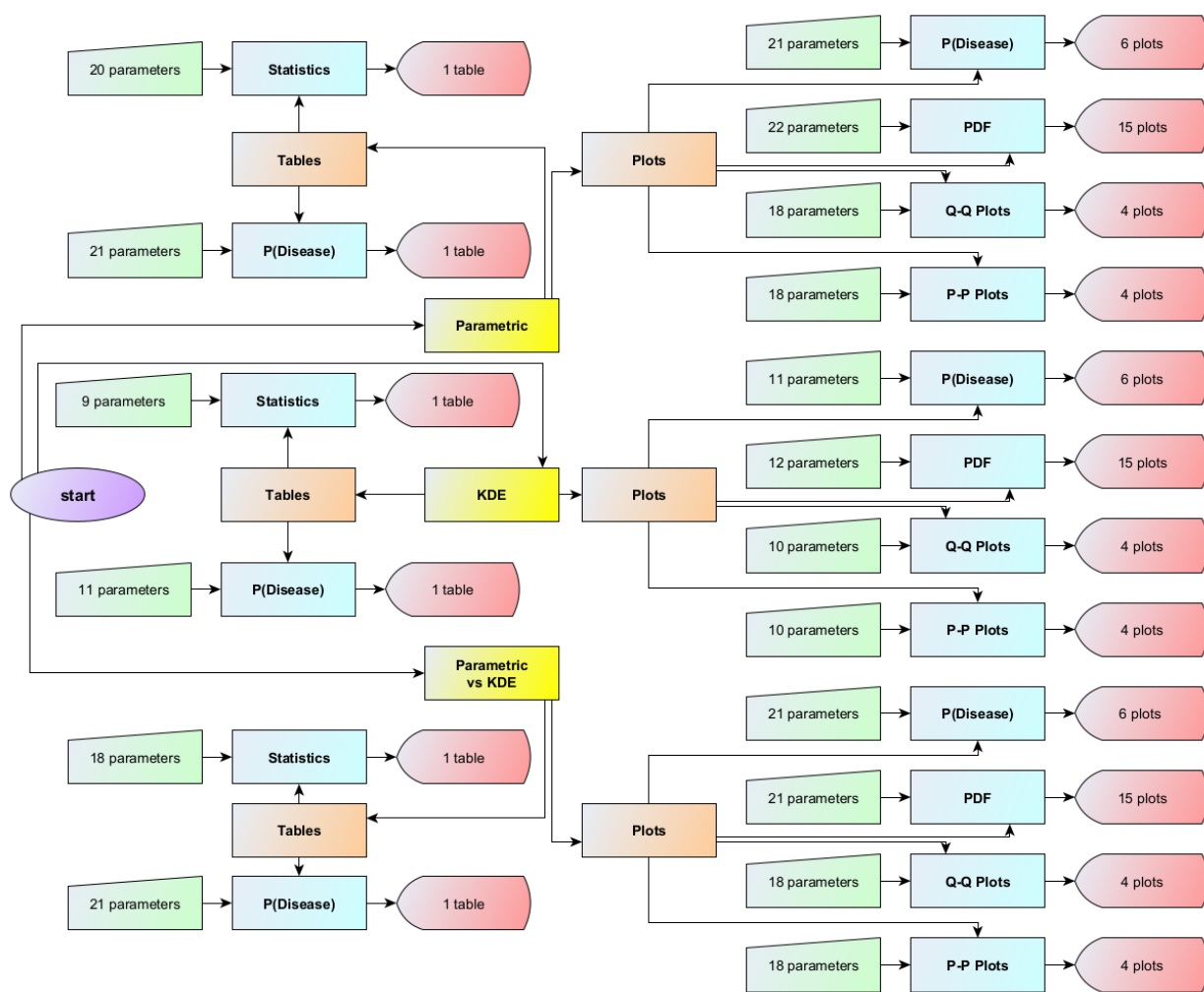


Figure 1. The flowchart of the Bayesian Diagnosis program.

2.2.2. Parametric Distributions

Parametric statistics assume that dataset data comes from a population that can be adequately modeled with a probability distribution that has a fixed set of parameters [19]. The parametric distributions provided by the program are the following:

1. Normal Distribution
 - 1.1 Univariate
 - 1.2 Bivariate
2. Lognormal Distribution
 - 2.1 Univariate
 - 2.2 Bivariate
3. Gamma Distribution
 - 3.1 Univariate
 - 3.2 Bivariate
4. Copula Distributions

The copula distributions of the program are bivariate, with a bivariate normal distribution with correlation ρ as kernel, and univariate normal, lognormal and gamma marginals.

The probability density functions (PDFs) of the parametric distributions are mathematically defined in Appendix A.

2.2.3. Nonparametric Distributions

Conversely, nonparametric models were also employed, which do not make a priori assumptions about the distribution's mathematical form [20]. These are particularly useful for exploratory data analysis and are implemented as shown in Appendix A.

Histograms

A histogram is the graphical representation of the distribution of a dataset as a series of bins.

The program plots histograms of the provided datasets.

Kernel Density Estimators (KDEs)

In contrast to histograms, a KDE generates a continuous and smooth estimate of the underlying PDF by summing the contributions of kernel functions centered at each data point.

KDEs offer a flexible nonparametric approach to density estimation, allowing for a better representation of the underlying distribution of the data.

The program provides univariate and bivariate Gaussian KDEs. The bivariate KDEs use radial-type kernels.

2.3. Interface of the Program

The *Bayesian Diagnosis* program is equipped with an intuitive tabbed user interface (refer to Figure 2). This design facilitates seamless navigation through its various modules and submodules. Users have the capability to input and modify a range of parameters, including prior probabilities and measurement parameters. Additionally, the interface allows for the selection of both parametric distributions and KDEs pertinent to medical diagnosis (refer to Appendix C and Supplementary File S1).

2.3.1. Input Parameters

Prior Probability

The user initiates the diagnostic evaluation by specifying the prior probability of disease occurrence in the population under study. This serves as a foundational measure for subsequent analyses.

Parametric Distributions

To facilitate a diagnostic model, the program allows for the definition of various parametric distributions for both the diseased and nondiseased populations across two diagnostic tests.

1. Distribution Selection: The user selects the type of distribution from a predefined list:
 - 1.1 Normal Distribution.
 - 1.2 Lognormal Distribution.
 - 1.3 Gamma Distribution.
2. Statistical Parameters: For each chosen distribution, the user defines the mean μ and standard deviation σ of the measurand in the respective population.
3. Correlation Coefficients: The user specifies the correlation coefficients ρ between the measurands of the first and second diagnostic tests for both diseased and nondiseased populations.

KDEs

Alternatively, the user can opt to define the KDEs for the measurands in both diseased and nondiseased populations across the two tests:

1. Bandwidth Parameter: For each KDE, the user defines the bandwidth parameter h .
2. Correlation Coefficients: As with parametric distributions, the user defines the correlation coefficients ρ between the measurands of the two diagnostic tests.

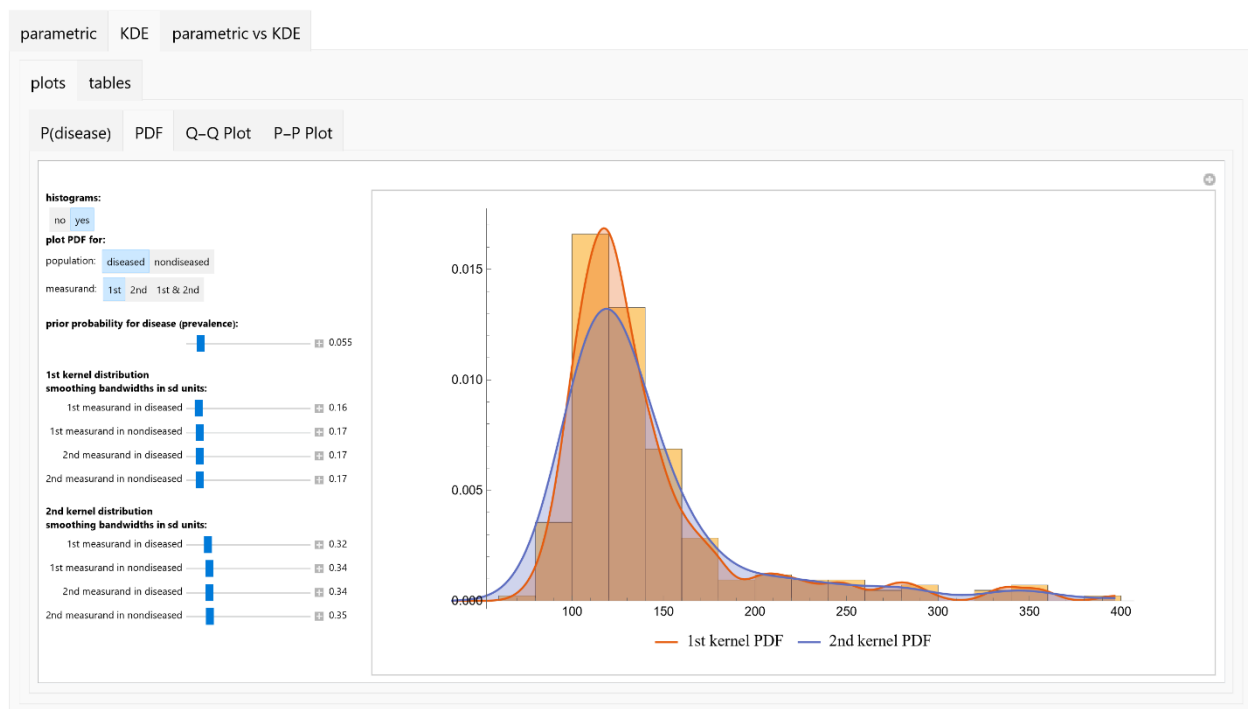


Figure 2. Screen shot of the *Bayesian Diagnosis* program.

2.3.2. Output Specifications

Visualizations

The program generates a series of plots designed to elucidate various diagnostic metrics and statistics:

1. **Posterior Probability of Disease:** Plots are generated to show the posterior probability of disease for each measurand and their combination.
2. **PDFs:** Univariate PDFs for each measurand and the bivariate PDF of their combination are plotted. An option to overlay histograms on these plots is also provided.
3. **Quantile–Quantile (Q–Q) Plots:** These plots are produced for each measurand to examine its distributional characteristics [21].
4. **Probability–Probability (P–P) Plots:** Similar to Q–Q plots, P–P plots are generated for further assessment of the distribution of each measurand [21].

The descriptions of the Q–Q and P–P plots are presented in the Supplementary File S2.

Tables

1. **Population Statistics:** The program tabulates key statistical metrics such as mean, median, standard deviation, skewness, kurtosis, and prior probability for each user-defined distribution and dataset. For each bivariate distribution of the two measurands in diseased and nondiseased populations, the correlation coefficients are calculated and displayed.
2. **Posterior Disease Probabilities:** For a user-defined pair of test measurement values, the program computes and presents the posterior probabilities for disease for each measurand and their combination.

By providing this comprehensive set of input parameters and output specifications, the program offers a robust platform for exploring the Bayesian diagnosis of disease using either parametric distributions or KDEs of medical diagnostic measurands.

2.3.3. Illustrative Application

To demonstrate the application of the program, fasting plasma glucose (FPG) was used as the first measurand and glycated hemoglobin A1c (HbA1c) as the second measurand for

Bayesian diagnosis of diabetes mellitus. The oral glucose tolerance test (OGTT) was used as the reference diagnostic method. A diagnosis of diabetes was confirmed if the plasma glucose (PG) value was equal to or exceeded 200 mg/dL, measured two hours after oral administration of 75 g of glucose [22], during an OGTT (2-h PG). It is noteworthy that the study population was confined to individuals aged between 40 and 60 years, a decision informed by the well-documented strong correlation between age and the prevalence of diabetes [23].

National Health and Nutrition Examination Survey (NHANES) data from participants was retrieved for the period from 2005 to 2016 [24] ($n = 60,936$). NHANES is a series of studies designed to evaluate the health and nutritional status of adults and children in the United States.

The inclusion criteria for participants were:

1. Age 40–60 years ($n = 11,782$);
2. Valid FPG, HbA1c, and OGTT measurements ($n = 4015$);
3. A negative response to NHANES question DIQ010 regarding a diabetes diagnosis [25] ($n = 3854$);
4. Non-pregnancy status ($n = 3854$).

Participants with a 2-h PG measurement of ≥ 200 mg/dL were considered diabetic ($n = 211$).

Descriptive statistics, including the mean, median, and standard deviation, were computed for each dataset and correlation coefficients for their combination. Univariate distributions were employed to model the distributions of FPG and HbA1c and bivariate distributions to model the joint distribution of FPG and HbA1c. Likelihoods and posterior probabilities were estimated for FPG, HbA1c and their combination.

The prior probability of diabetes was estimated as follows:

$$v = \frac{211}{3854} = 0.055$$

The statistics of the dataset are presented in Table 1.

Table 1. The descriptive statistics of FPG and HbA1c datasets.

	Diabetic Patients		Nondiabetic Patients	
	<i>n</i>	687	10,519	
Measurand (Units)	FPG (mg/dL)	HbA1c (%)	FPG (mg/dL)	HbA1c (%)
Mean	141.3	6.67	99.9	5.47
Median	124.0	6.30	99.0	5.50
Standard Deviation	54.0	1.57	10.1	0.38
Skewness	2.375	2.201	0.576	−0.058
Kurtosis	9.037	8.377	4.213	3.615
Correlation Coefficient	0.914		0.320	

3. Results

Using the settings of Table 2, the program generated the plots of Figures 3–13 and the tables of Figures 14 and 15.

Table 2. The settings of the program for Figures 3–15.

	Diabetic Patients		Nondiabetic Patients	
	FPG (mg/dL)	HbA1c (%)	FPG (mg/dL)	HbA1c (%)
Measurand (Units)				
Parametric Distribution	Lognormal	Lognormal	Lognormal	Lognormal
Parametric Distribution Mean	141.3	6.67	99.9	5.47
Parametric Distribution SD	54.0	1.57	10.1	0.38
KDE Smoothing Bandwidth (SD units)	0.32	0.34	0.34	0.35
Correlation Coefficient	0.914		0.320	

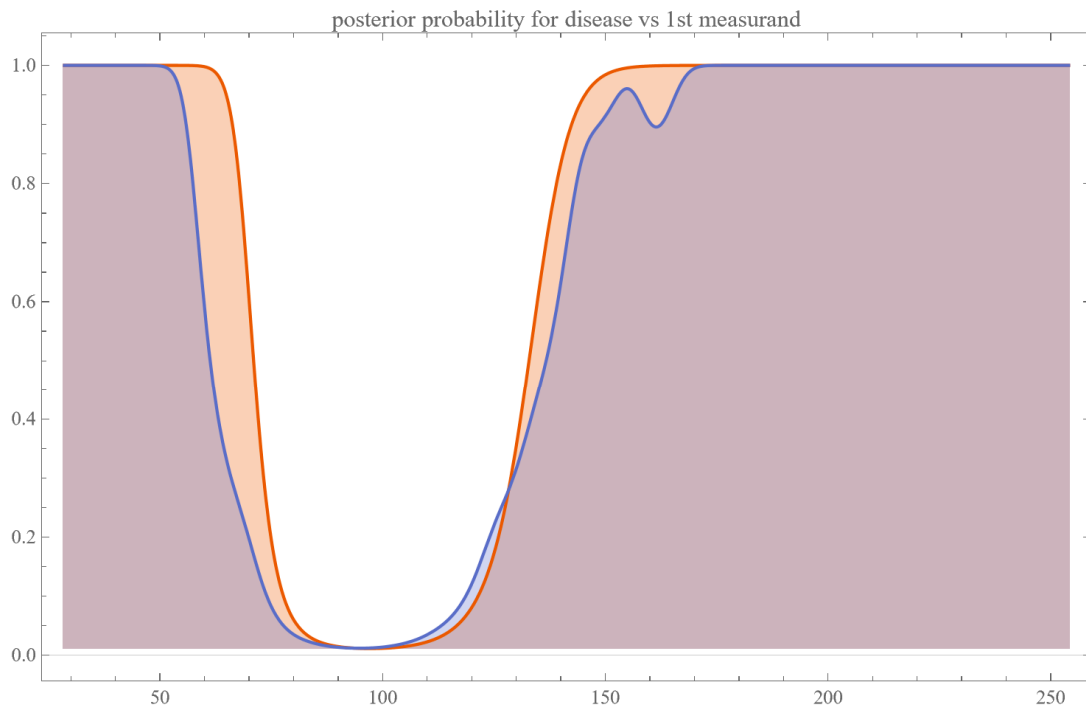


Figure 3. Posterior probability of disease (diabetes) versus the first measurand (FPG), assuming parametric and KDE distributions of the measurand, with the settings of the program in Table 2.

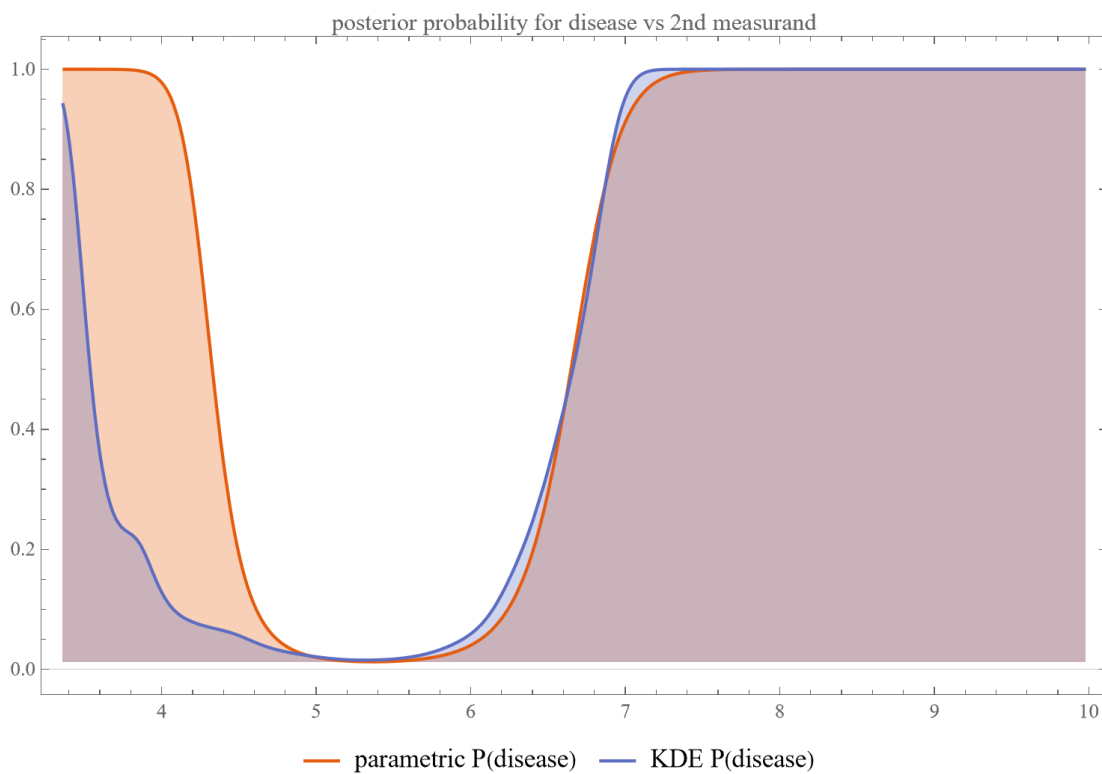


Figure 4. Posterior probability of disease (diabetes) versus the second measurand (HbA1c), assuming parametric and KDE distributions of the measurand, with the settings of the program in Table 2.

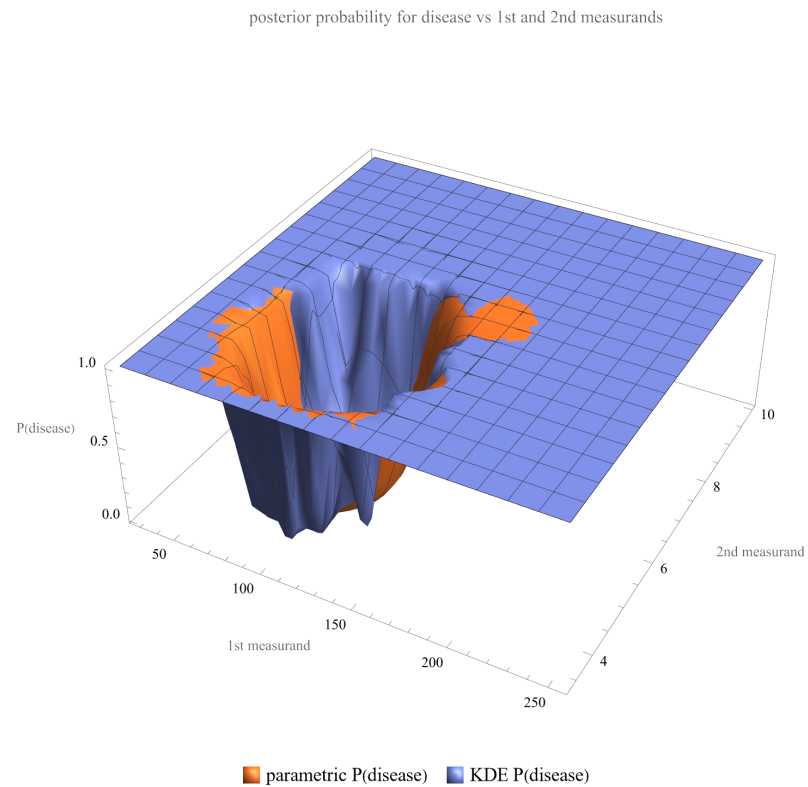


Figure 5. Posterior probability of disease (diabetes) versus both measurands (FPG and HbA1c), assuming parametric and KDE distributions of the measurands, with the settings of the program in Table 2.

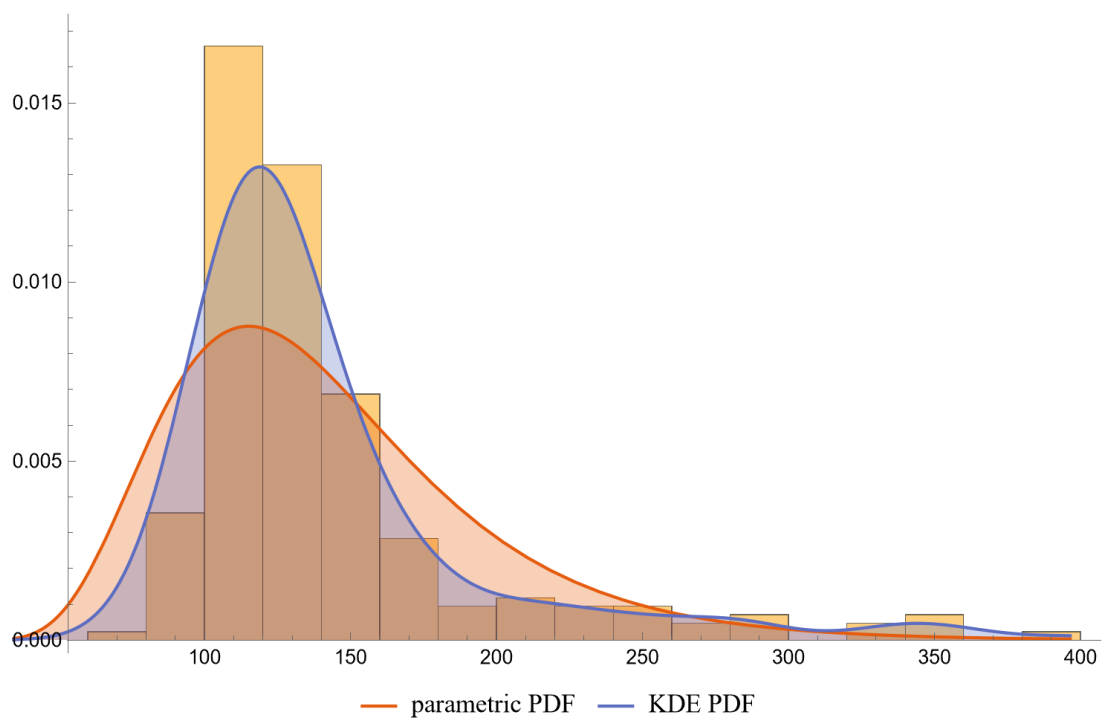


Figure 6. The PDF of the first measurand (FPG) in diseased (diabetic patients), assuming parametric and KDE distributions of the measurand, and the histogram of the respective dataset (NHANES dataset), with the settings of the program in Table 2.

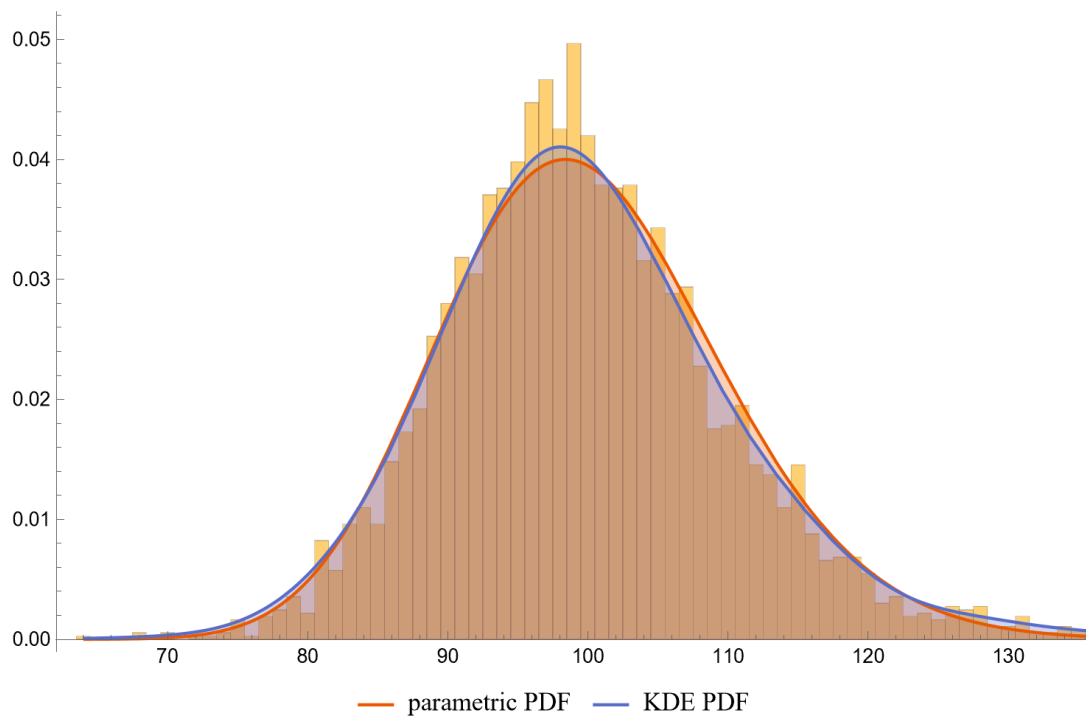


Figure 7. The PDF of the first measurand (FPG) in nondiseased (nondiabetic patients), assuming parametric and KDE distributions of the measurand, and the histogram of the respective dataset (NHANES dataset), with the settings of the program in Table 2.

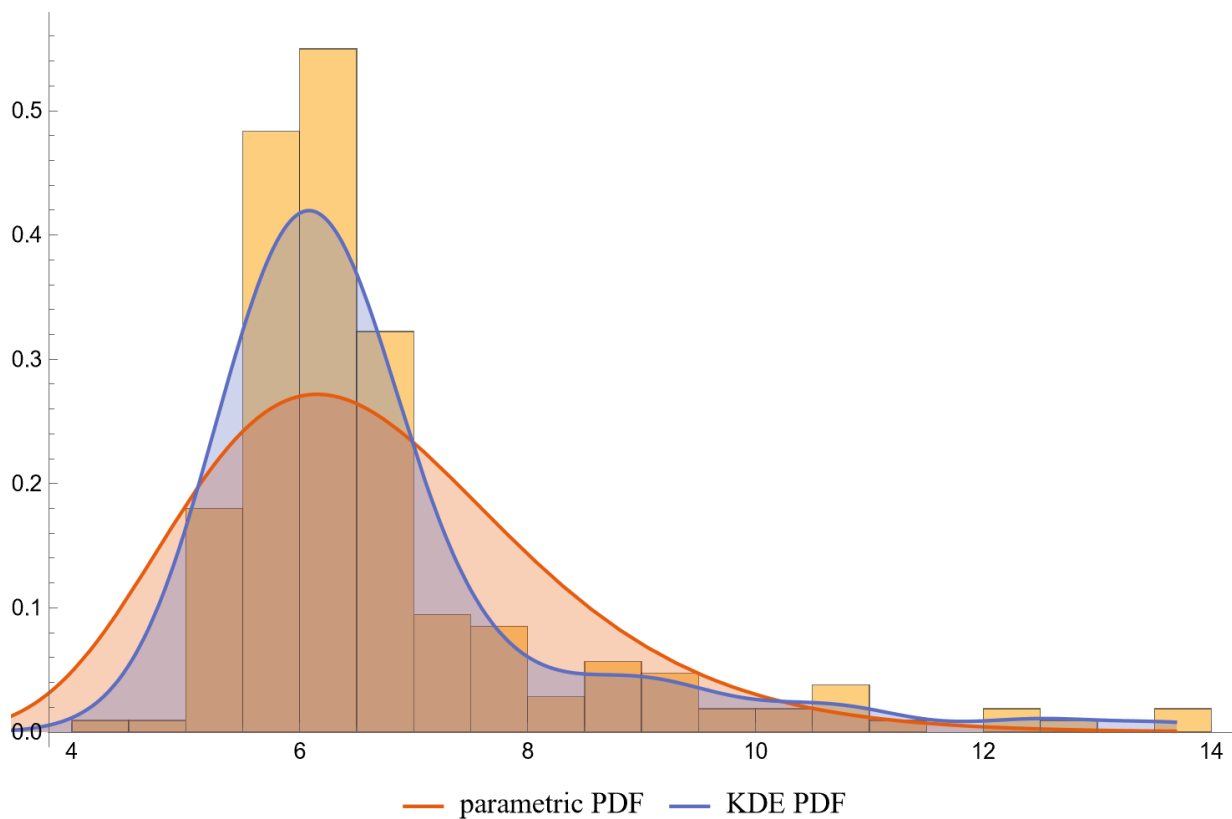


Figure 8. The PDF of the second measurand (HbA1c) in diseased (diabetic patients), assuming parametric and KDE distributions of the measurand, and the histogram of the respective dataset (NHANES dataset), with the settings of the program in Table 2.

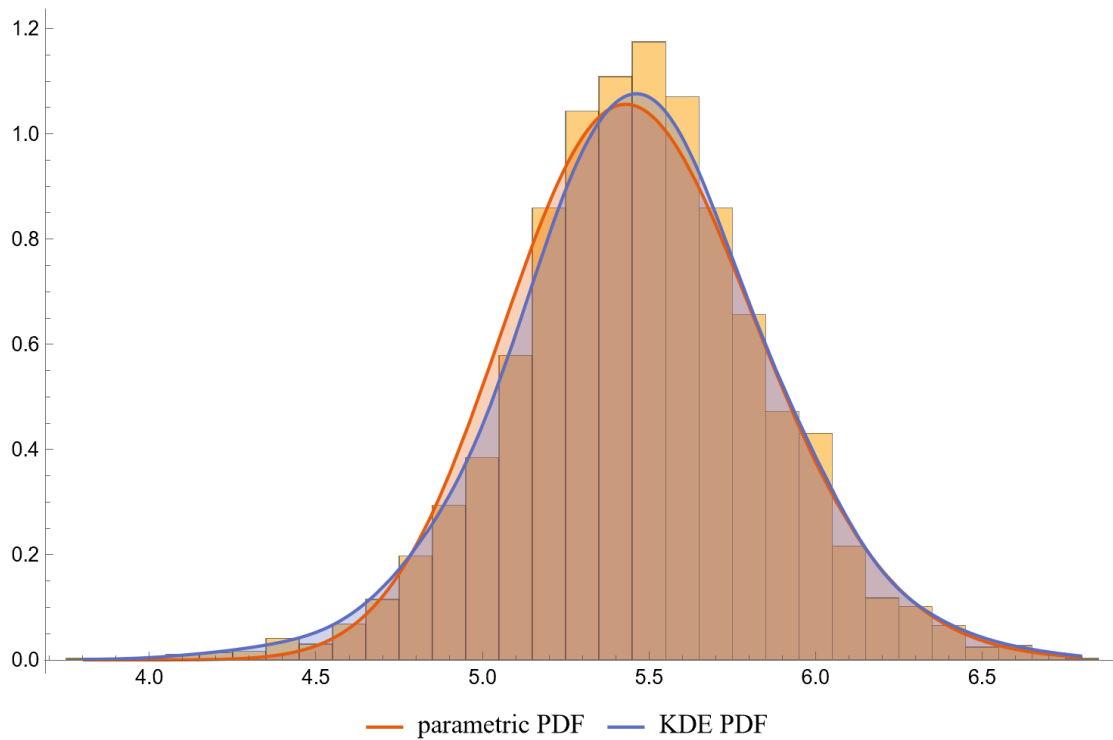


Figure 9. The PDF of the second measurand (HbA1c) in nondiseased (nondiabetic patients), assuming parametric and KDE distributions of the measurand, and the histogram of the respective dataset (NHANES dataset), with the settings of the program in Table 2.

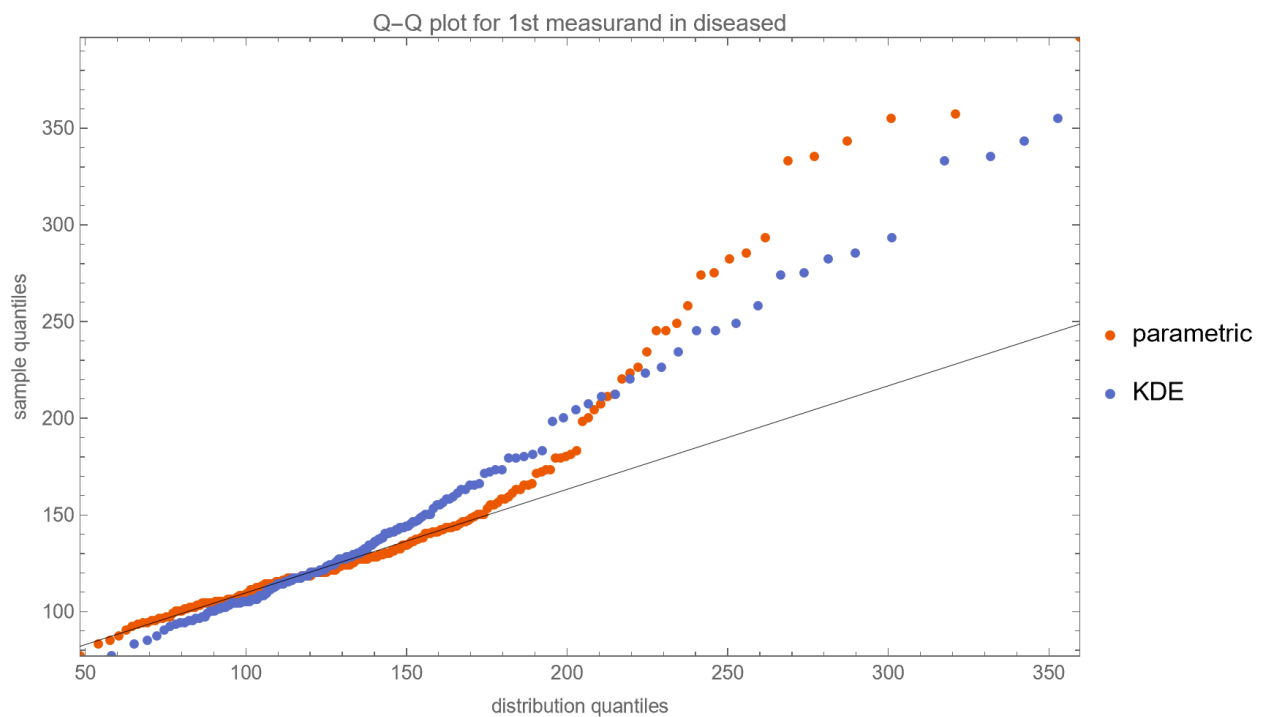


Figure 10. The Q–Q plot of the first measurand (FPG) in diseased (diabetic patients) versus the respective dataset (NHANES dataset), assuming parametric and KDE distributions of the measurand, with the settings of the program in Table 2.

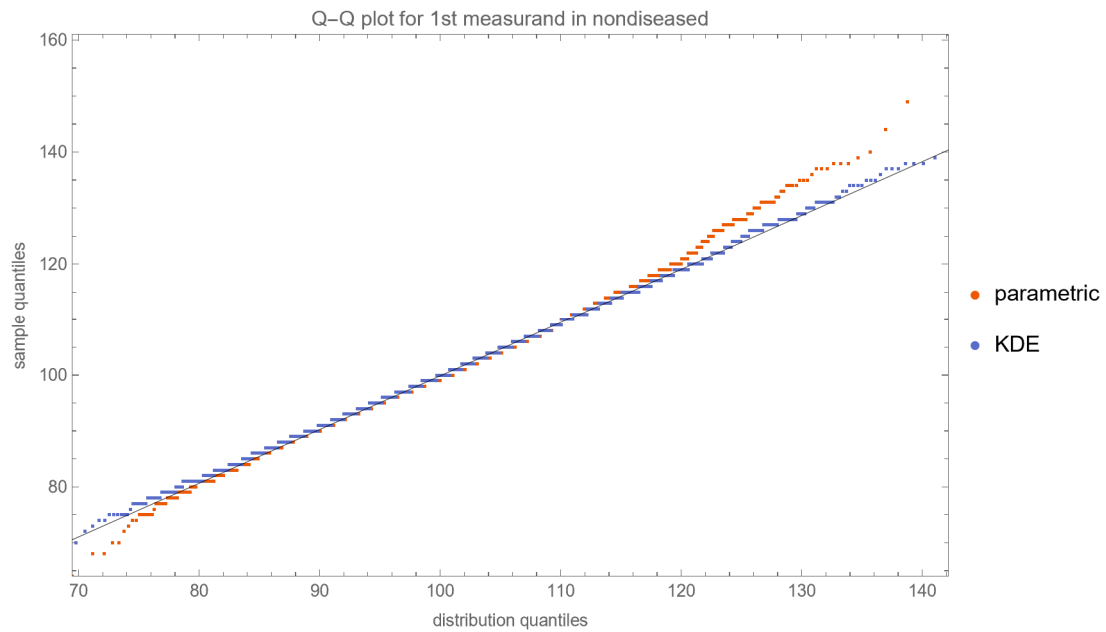


Figure 11. The Q–Q plot of the first measurand (FPG) in nondiseased (nondiabetic patients) versus the respective dataset (NHANES dataset), assuming parametric and KDE distributions of the measurand, with the settings of the program in Table 2.

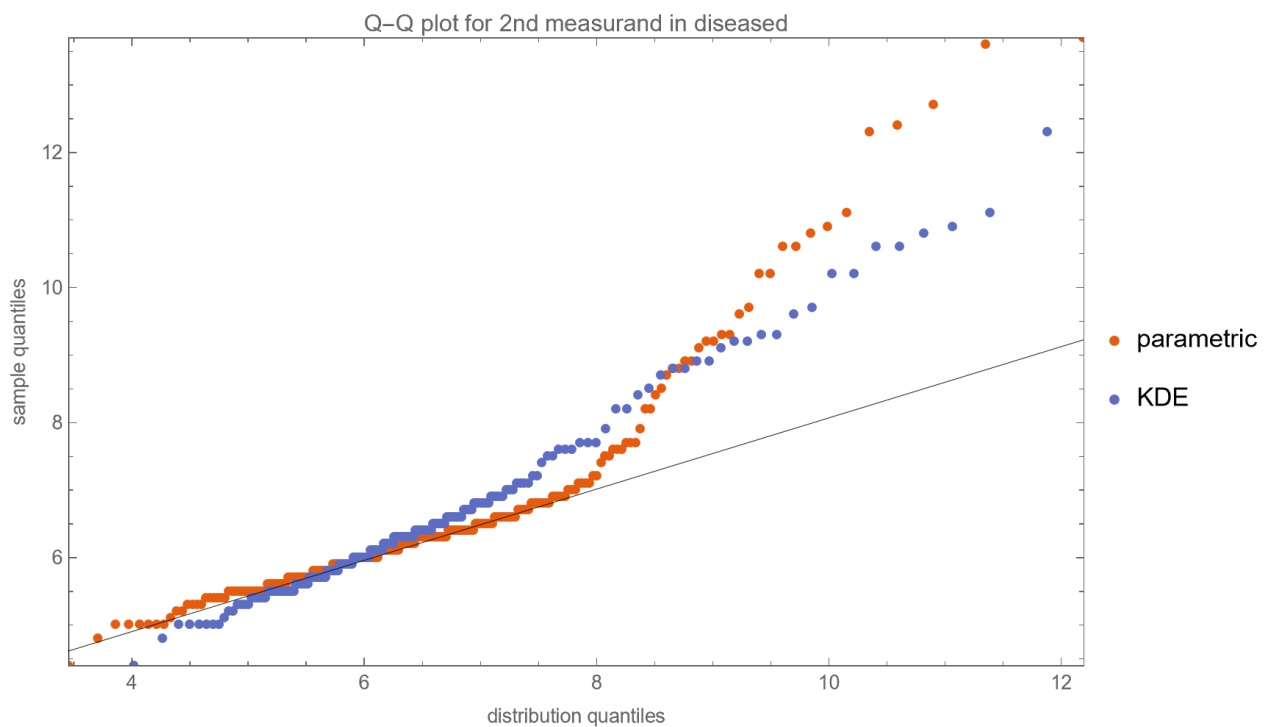


Figure 12. The Q–Q plot of the second measurand (HbA1c) in diseased (diabetic patients) versus the respective dataset (NHANES dataset), assuming parametric and KDE distributions of the measurand, with the settings of the program in Table 2.

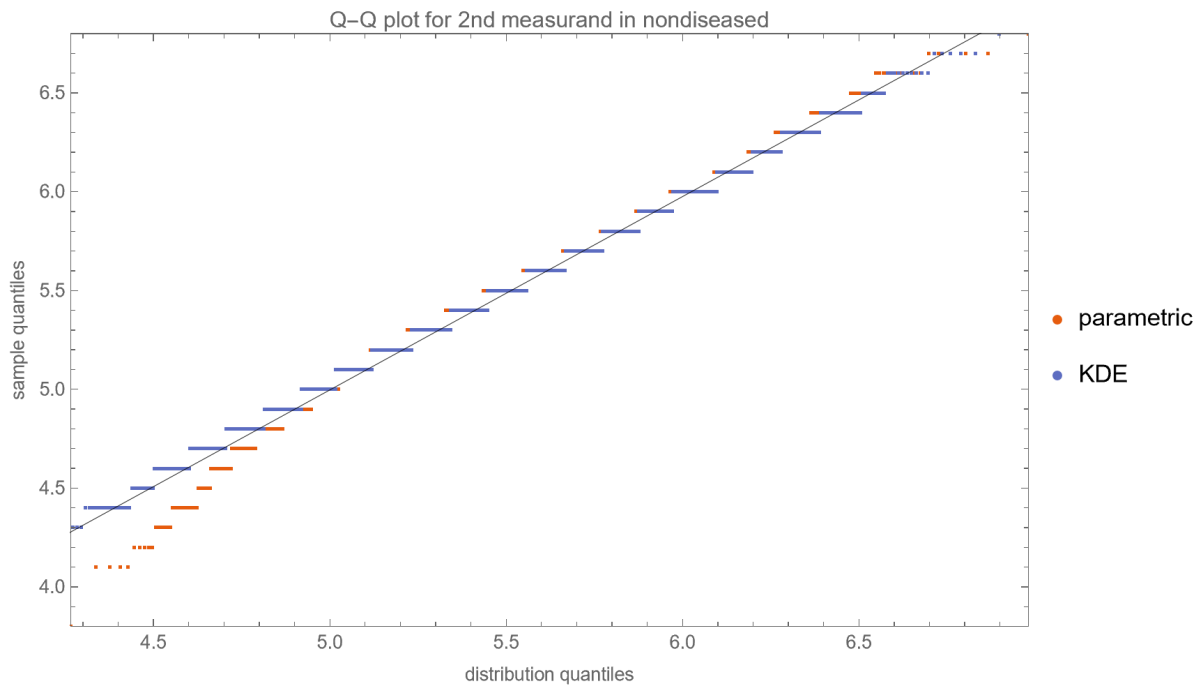


Figure 13. The Q–Q plot of the second measurand (HbA1c) in nondiseased (nondiabetic patients) versus the respective dataset (NHANES dataset), assuming parametric and KDE distributions of the measurand, with the settings of the program in Table 2.

measurements statistics							
		diseased			nondiseased		
	parameters	parametric	KDE	dataset	parametric	KDE	dataset
1st measurand	mean	141.3000	141.2512	141.2512	99.9000	99.8636	99.8636
	median	131.9898	127.2044	124.0000	99.3933	99.1482	99.0000
	sd	54.0000	56.5278	53.9507	10.1000	10.6961	10.1265
	skewness	1.2023	2.0498	2.3746	0.3043	0.4888	0.5763
	kurtosis	5.6759	7.9620	9.0374	3.1651	3.9743	4.2133
	log-likelihood	-1086.9440	-1045.4940		-13537.3200	-13518.6800	
	prior probability	0.055	0.055	0.055	0.945	0.945	0.945
	correlation coefficient	0.9140	0.8141	0.9139	0.3200	0.2791	0.3204
2nd measurand	mean	6.6700	6.6697	6.6697	5.4700	5.4705	5.4705
	median	6.4926	6.2863	6.3000	5.4568	5.4693	5.5000
	sd	1.5700	1.6552	1.5704	0.3800	0.3996	0.3772
	skewness	0.7192	1.8660	2.2009	0.2087	-0.0489	-0.0582
	kurtosis	3.9336	7.3146	8.3766	3.0776	3.4876	3.6147
	log-likelihood	-359.0682	-313.1941		-1643.1330	-1599.1410	
	prior probability	0.055	0.055	0.055	0.945	0.945	0.945

Figure 14. Descriptive statistics of the distributions of the measurands (FPG and HbA1c) in diseased (diabetic patients) and nondiseased (nondiabetic patients), assuming parametric and KDE distributions, and of the respective datasets (NHANES datasets), with the settings of the program in Table 2.

prior probability for disease			
prevalence	0.055		
posterior probability for disease			
	parametric	KDE	difference
1st measurand = 126	0.200471	0.238737	0.0382659
2nd measurand = 6.5	0.294458	0.331717	0.0372596
1st measurand = 126 & 2nd measurand = 6.5	0.480279	0.344877	-0.135402

Figure 15. The prior and posterior probabilities of disease (diabetes) for values of the first measurand (FPG) equal to 126 mg/dL and of the second measurand (HbA1c) equal to 6.5%, assuming parametric and KDE distributions, with the settings of the program in Table 2.

The KDEs smoothing bandwidth was set to double that given with Silverman's rule of thumb [26,27].

Figures 3 and 4 show the plots of the posterior probability of diabetes versus FPG and HbA1c, respectively. The curves of the parametric distributions are smooth double sigmoidal, while the curves of the nonparametric distributions are multimodal.

Figure 5 shows the plot of the posterior probability of diabetes versus FPG and HbA1c combined. The surface of the parametric distribution is smooth, while the surface of the nonparametric distribution is multimodal.

Figures 6–9 show the PDF of FPG and HbA1c in diabetic and nondiabetic patients and the histograms of the respective NHANES datasets. It is visually evident that the nonparametric distributions fit the datasets better, especially in diabetic patients.

Figures 10–13 show the Q–Q plots of the parametric and KDE distributions of FPG and HbA1c in diabetic and nondiabetic patients versus the respective NHANES datasets. The plots show clearly that the nonparametric distributions fit the datasets better, especially in diabetic patients.

Figure 14 shows a table with the descriptive statistics of FPG and HbA1c in diabetic patients and nondiabetic patients, assuming parametric and KDE distributions, and of the respective NHANES datasets. The data, including the loglikelihood values, supports the hypothesis that the nonparametric distributions fit the datasets better, especially in diabetic patients.

Figure 15 shows a table of prior and posterior probabilities for disease (diabetes) for values of FPG equal to 126 mg/dL and of HbA1c equal to 6.5%, the established thresholds of the two measurands for the diagnosis of diabetes [22], assuming parametric and KDE distributions.

4. Discussion

4.1. Reevaluation of Traditional Diagnostic Methods

The findings of the present study highlight the importance of considering incorporating Bayesian methods in medical diagnosis and management. Conventional approaches based on rigid diagnostic criteria are often unable to account for the intricate relationships between disease pathology and diagnostic procedures thus limiting personalized patient care options. [28]. In stark contrast, Bayesian methodologies offer a framework that enhances diagnostic precision through a more comprehensive probabilistic assessment [5]. This Bayesian foundation, therefore, could serve as an enabler for tailored medical interventions, echoing similar arguments in existing literature advocating for individualized medicine [29].

The study population was confined to individuals aged between 40 and 60 years. This restriction allowed for a more homogeneous prior probability, thereby reducing the impact of age-specific variations in the prevalence of the condition under study.

Even though the KDEs from our illustrative application, as parameterized in Table 2, provide only an approximate fit to the NHANES datasets for FPG and HbA1c measurements, the posterior probabilities for diabetes delineated in Figure 15 suggest a limited concordance between the classification criteria of diabetes derived from the OGTT, HbA1c, and FPG tests [22], as found previously in existing literature [30].

4.2. Challenges and Considerations in Bayesian Analysis for Disease Diagnosis

Despite the evident merits of Bayesian analytics in medical diagnostics, it is paramount to address the intrinsic challenges associated with this methodological shift. One such issue resides in the limited availability of scholarly publications that provide a comprehensive statistical exploration of the measurands in both the diseased and nondiseased populations [31].

4.2.1. Ramifications of Incomplete Information

1. **Over-dependence on Prior Probabilities:** The scarcity of empirically derived distributions amplifies reliance on prior probabilities, thereby inducing distortions in the calculation of posterior probabilities. This could result in suboptimal clinical judgments and potentially inaccurate diagnoses [32].
2. **Elevated Uncertainty:** Insufficient data contributes to broader confidence intervals in the computed posterior probabilities, which, in turn, could exacerbate clinical indecisiveness [33].
3. **Risk of Bias:** The introduction of systemic bias due to unrepresentative datasets could compromise the fidelity of Bayesian calculations [7].
4. **Imperative for Collaborative Research:** More coordinated research is needed, including multi-center studies, meta-analyses, and open-access databases—to accumulate and disseminate data essential for effective Bayesian diagnosis [34].
5. **Exploration of Alternative Methodologies:** Given the lack of comprehensive data, the utility of combining Bayesian methods with other statistical and computational techniques or diagnostic modalities becomes increasingly pertinent [35,36].

4.2.2. Parametric Versus Nonparametric Bayesian Models

In the context of diagnosing diabetes mellitus through FPG and HbA1c levels, our computational tool revealed that nonparametric Bayesian models typically produce a better fit to data distributions, corroborating existing literature that emphasizes the robustness of nonparametric techniques in capturing complex data distributions [26,37].

4.2.3. Multimodal Versus Double Sigmoidal Bayesian Probability of Disease Curve

The nonparametric Bayesian probabilities for disease exhibited multimodal patterns, in contrast to the bimodal, double sigmoidal curves generated by parametric models.

Multimodal Curve

Potential Causes:

- (a) **Complex Pathophysiology:** Multiple etiological pathways may influence the same measurand in divergent ranges, adding layers of complexity to diagnostic processes [13].
- (b) **Diagnostic Confounders:** External variables affecting the measurand could compromise its efficacy as a standalone diagnostic criterion [38].
- (c) **Population Subgroups:** The existence of demographically or genetically distinct subgroups within the studied population could also account for the observed multimodality [39].

- (d) **Statistical Artifacts:** Demographically or genetically distinct subgroups may be a factor contributing to observed multimodal distributions [39].

Theoretical Implications:

Multimodal distributions present a clinical conundrum, compelling healthcare providers to potentially employ additional diagnostic tests or even alternative methodologies [13].

Double Sigmoidal Curve

A curve composed of two mirrored sigmoid functions, one delineating the probability behavior for lower measurand values and the other for higher values—presents a compelling intricacy in the realm of diagnostic statistics and medical decision-making.

Interpretation

- (a) **Two Zones of Risk:** Such a curve suggests that the risk of the disease is heightened both at low and high extremes of the measurand but reduced in a middle “safe zone.”
- (b) **Multifactorial Etiology:** This might reflect a situation where both deficiency and excess of a particular biological factor contribute to disease risk. For example, both low and elevated levels of hormones may pose challenges to physiological homeostasis.

Clinical and Diagnostic Implications

- (a) **Threshold Decision-making:** Unlike a single sigmoid curve, where one threshold may be adequate for diagnosis, the double sigmoid may necessitate multiple thresholds, defining a “safe zone” for the measurand.
- (b) **Treatment Strategies:** Clinicians must be cautious when intervening based on such a measurand, as moving the measurand too far in either direction could heighten risk.
- (c) **Population Stratification:** This curve shape might imply that different sub-populations or disease subtypes could be better distinguished with additional tests or measurements.

4.3. Shortcomings of This Study

The main shortcomings of this study were the following:

1. The OGTT was used as a reference diagnostic method for diabetes mellitus. The diagnostic threshold for 2-h PG was established in relation to the risk of diabetic retinopathy, a microvascular complication of diabetes mellitus [40]. However, glucose tolerance is influenced by complex interactions of factors, both physiological and environmental, which pose significant implications for clinical diagnosis and research. The considerations that could affect glucose tolerance and, therefore, the interpretation of the 2-h PG measurement, include the following:
 - (a) **Age and Gender:** Age and gender are significant variables in glucose tolerance. Insulin sensitivity often decreases with age, resulting in higher 2-h PG levels [41]. Gender differences, particularly related to hormonal changes in females, could also affect glucose metabolism [42].
 - (b) **Diurnal Variability:** Glucose tolerance is subject to diurnal variation, which could affect the 2-h PG test outcomes. Insulin sensitivity is generally higher in the morning than in the evening [43].
 - (c) **Physical Activity:** Exercise improves insulin sensitivity and therefore could affect glucose tolerance tests. The timing and intensity of physical activity could have a direct influence on the 2-h PG results [44].
 - (d) **Dietary Patterns:** Short-term and long-term dietary habits, including the macronutrient composition of the diet, may alter the body’s glucose and insulin response [45].
 - (e) **Stress and Emotional States:** The acute stress response includes a transient rise in glucose levels as a result of catecholamine release, potentially affecting the 2-h PG test [46].
 - (f) **Medications:** Certain medications like corticosteroids, antipsychotics, and diuretics affect glucose metabolism, thereby influencing 2-h PG test outcomes [47].

- (g) Genetic Factors: Genetic predispositions influence glucose tolerance, and not accounting for this introduce variability in the 2-h PG test [48].
2. The lognormal distributions and the KDE, as parameterized in Table 2, fitted only approximately to the NHANES datasets of FPG and HbA1c measurements. It is well known that biological measurands, such as FPG and HbA1c, may not follow textbook statistical distributions like normal or lognormal distributions. Numerous papers have noted the skewness or kurtosis in the distribution of metabolic variables, urging the use of flexible statistical models [49,50].

Related Statistical Software

All major general or Bayesian statistical software packages (OpenBUGS, Ver. 3.2.3, JASP[®], Ver. 0.18.1, Matlab[®], Ver. R2023b, NCSS[®], Ver. 23.0.2, R, Ver. 4.3.1, SAS[®], Ver.9.4M8, SPSS[®], Ver. 29, Stan, Ver. 2.33.0, and Stata[®] Ver. 18) include routines for Bayesian inference. The program presented in this work provides 29 different types of parametric and nonparametric plots. None of the above-mentioned programs provide this range of plots without advanced statistical programming.

5. Conclusions and Future Directions

The intricacies of the double-sigmoid curve and multimodal distributions introduce a new frontier in personalizing healthcare provision. While smoother relationships between measurements and Bayesian probability facilitate clinical interpretability, multimodal distributions might serve as sentinel indicators of underlying complexities or methodological shortcomings, thus providing a useful tool in the field of medical diagnosis.

As a pivotal next step, future research should aim to validate the utility and reliability of the Bayesian inference based method applied in this study through real-world clinical trials, in addition to extending its application to include more diagnostic modalities. The aim is to combine this approach with existing clinical protocols, thereby optimizing the diagnostic precision and consequently improving patient outcomes.

In addition to its potential for clinical applications, the computational tool developed for this study could hold considerable promise as an educational and research adjunct. By facilitating the analysis of Bayesian probabilities in disease diagnosis, it serves as an invaluable resource for both medical practitioners in training and experienced researchers in the field. Its modular design and user-friendly interface make it easily adaptable to various research settings and educational curricula, thereby accelerating the adoption and dissemination of Bayesian approaches in medical statistics and diagnostics.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/diagnostics13193135/s1>, BayesianDiagnosis.nb: The program as a Wolfram Notebook. Available at <https://www.hcsl.com/Tools/BayesianDiagnosis/BayesianDiagnosis.nb> (accessed on 28 September 2023); Supplementary File: BayesianDiagnosis.nb can be found at <https://zenodo.org/record/8414309> (accessed on 6 October 2023); Supplementary File S1 can be found at <https://zenodo.org/record/8407804> (accessed on 6 October 2023); Supplementary File S2 can be found at <https://zenodo.org/record/8414841> (accessed on 6 October 2023).

Author Contributions: Conceptualization: T.C.; methodology: T.C. and A.T.H.; software: T.C. and A.T.H.; validation: T.C.; formal analysis: T.C. and A.T.H.; investigation: T.C.; resources: A.T.H.; data curation: T.C.; writing—original draft preparation: T.C.; writing—review and editing A.T.H.; visualization: T.C.; supervision: A.T.H.; project administration: T.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Data collection was carried out following the rules of the Declaration of Helsinki. The Ethics Review Board of the National Center for Health Statistics approved data collection and posting the data online for public use [51] (National Center for Health Statistics 2022).

Informed Consent Statement: Written consent was obtained from each subject participating in the survey.

Data Availability Statement: The data presented in this study is available at <https://wwwn.cdc.gov/nchs/nhanes/default.aspx> (accessed on 4 September 2023).

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

PDF	probability density function
CDF	cumulative distribution function
KDE	kernel density estimator
OGTT	oral glucose tolerance test
PG	plasma glucose
2-h PG	plasma glucose, measured two hours after oral administration of 75 g of glucose, during an OGTT
FPG	fasting plasma glucose
HbA1c	glycated hemoglobin A1c
NHANES	National Health and Nutrition Examination Survey

Appendix A

Appendix A.1. Formalisms and Notation

Appendix A.1.1. Tuples

\mathbf{x} : an n -tuple (x_1, x_2, \dots, x_n)

Appendix A.1.2. Parameters

v : prevalence of disease

μ, m : mean

σ, s : standard deviation

ρ, r : correlation coefficient

k : shape parameter

θ : scale parameter

h : nonparametric kernel density bandwidth

Appendix A.1.3. Functions

f^{-1} : the inverse of the function f

$|H|$: determinant of the matrix H

$P(A)$: probability of the event A

$P(A, B)$: conditional probability of the event A given the event B

$cov(X, Y)$: covariance of two jointly distributed random variables X and Y

$\mathbb{E}[Z]$: expected value of a random variable Z

$\ln(x)$: natural logarithm

$\mathcal{L}(\theta|z)$: likelihood function of the parameter θ given the observed value z of the random variable Z

$\mathcal{L}(\theta|\mathbf{z})$: likelihood function of the parameter θ given the observed values \mathbf{z} of the random variable Z .

$l(\theta|z)$: loglikelihood function of the parameter θ given the observed value z of the random variable Z

$l(\theta|\mathbf{z})$: loglikelihood function of the parameter θ given the of observed values \mathbf{z} of the random variable Z .

$p(x)$: probability mass function of a discrete variable X

$P_Q(k; q)$: the k^{th} q -quantile of a random variable

$erf(z)$: error function

$erfc(z)$: complementary error function

$\Gamma(z)$: gamma function
 $\Gamma(z, x)$: incomplete gamma function
 $Q(a, z)$: regularized incomplete gamma function
 $\Gamma(z, x_0, x_1)$: generalized incomplete gamma function
 $Q(z, x_0, x_1)$: regularized generalized incomplete gamma function
 $K(u)$: kernel function
 $f(x)$: univariate PDF
 $f(x; \theta), f(x|\theta)$: univariate PDF given the multivariate parameter θ
 $f(x, y)$: bivariate PDF
 $f(x, y; \theta), f(x, y|\theta)$: bivariate PDF given the multivariate parameter θ
 $F(x)$: univariate CDF
 $F(x; \theta), F(x|\theta)$: univariate CDF given the multivariate parameter θ
 $F(x, y)$: bivariate CDF
 $F(x, y; \theta), F(x, y|\theta)$: bivariate CDF given the multivariate parameter θ

Definitions of the aforementioned functions are presented in Supplementary File S2.

Appendix A.2. Bayes Theorem

For the purposes of our study, Bayes' theorem is formulated as follows:

$$P(D|T) = \frac{P(T|D)P(D)}{P(T)} = \frac{P(T|D)P(D)}{P(T|D)P(D) + P(T|\bar{D})(1 - P(D))}$$

where:

$P(D|T)$ represents the posterior probability of having the disease given the test results \mathbf{z} .
 $P(T|D)$ denotes the likelihood of obtaining the test results \mathbf{z} given the presence of the disease.

$P(T|\bar{D})$ denotes the likelihood of obtaining the test results \mathbf{z} given the absence of the disease.

$P(D)$ is the prior probability or prevalence v of the disease.

$P(T)$ signifies the overall probability of the test results \mathbf{z} .

Therefore, for the possibly multivariate parameter θ :

$$P(D|T) = \frac{\mathcal{L}_D(\theta|\mathbf{z})v}{\mathcal{L}_D(\mathbf{z}|\theta)v + \mathcal{L}_{\bar{D}}(\mathbf{z}|\theta)(1 - v)} = \frac{f_D(\mathbf{z}|\theta)v}{f_D(\mathbf{z}|\theta)v + f_{\bar{D}}(\mathbf{z}|\theta)(1 - v)}$$

where $\mathcal{L}_D(\theta|\mathbf{z})$ and $f_D(\mathbf{z}|\theta)$ denote the likelihood function and the PDF in the presence of the disease, while $\mathcal{L}_{\bar{D}}(\mathbf{z}|\theta)$ and $f_{\bar{D}}(\mathbf{z}|\theta)$ denote the respective functions in the absence of the disease.

Appendix A.3. Parametric Distributions

Appendix A.3.1. Normal Distribution

(a) Univariate

The univariate normal distribution or Gaussian distribution is a continuous probability distribution of a random variable X . The general form of its PDF is:

$$f_N(x; \mu, \sigma) = \frac{e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}{\sigma\sqrt{2\pi}}$$

where the parameter μ is the mean of X , while σ is its standard deviation [52].

(b) Bivariate

The bivariate normal distribution or Gaussian distribution is a continuous probability distribution of two normally distributed random variables X and Y . The general form of its PDF is:

$$f_N(x, y; \mu_X, \sigma_X, \mu_Y, \sigma_Y, \rho) = \frac{e^{-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_X)^2}{\sigma_X^2} - \frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \frac{(y-\mu_Y)^2}{\sigma_Y^2}\right)}}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}}$$

where μ_X and μ_Y are the means of X and Y , σ_X and σ_Y are their standard deviations, and ρ their correlation coefficient [52].

Appendix A.3.2. Lognormal Distribution

(a) Univariate

The univariate lognormal distribution is a continuous probability distribution of a random variable X whose logarithm is normally distributed. The general form of its PDF is:

$$f_L(x; m, s) = \frac{e^{-\frac{1}{2}\left(\frac{\ln(x)-m}{s}\right)^2}}{xs\sqrt{2\pi}}$$

where m is the mean and s the standard deviation of $\ln(X)$ [52].

If μ and σ are the mean and the standard deviation of X , we have:

$$\mu = e^{m+\frac{1}{2}s^2}$$

$$\sigma = \sqrt{e^{2m+2s^2}}$$

Therefore,

$$m = \ln\left(\frac{\mu^2}{\sqrt{\sigma^2 + \mu^2}}\right)$$

$$s = \ln\left(1 + \frac{\sigma^2}{\mu^2}\right)$$

$$f_L(x; \mu, \sigma) = \frac{e^{-\frac{1}{2}\left(\frac{\ln(x)-\ln\left(\frac{\mu^2}{\sqrt{\sigma^2+\mu^2}}\right)}{\ln\left(1+\frac{\sigma^2}{\mu^2}\right)}\right)^2}}{\sqrt{2\pi}x\ln\left(1+\frac{\sigma^2}{\mu^2}\right)} = \frac{e^{-\frac{\left(2\ln(x)-2\ln(\mu)+\ln\left(1+\frac{\sigma^2}{\mu^2}\right)\right)^2}{8\ln\left(1+\frac{\sigma^2}{\mu^2}\right)}}}{x\sqrt{2\pi\ln\left(1+\frac{\sigma^2}{\mu^2}\right)}}$$

(b) Bivariate

The bivariate lognormal distribution is a continuous probability distribution of two lognormally distributed variables X and Y . If m_X and m_Y are the means of $\ln(X)$ and $\ln(Y)$, s_X and s_Y their standard deviations, and r their correlation coefficient, the general form of its PDF is [52]:

$$f_L(x, y; m_X, s_X, m_Y, s_Y, r) = \frac{1}{d}e^a$$

where

$$a = \frac{1}{2}\left(\frac{-(\ln(y) - m_Y)b - (\ln(x) - m_X)c}{s_X^2\sigma_Y^2 - r^2s_X^2s_Y^2}\right)$$

$$b = (\ln(y) - m_Y)s_X^2 - r(\ln(x) - m_X)s_Xs_Y$$

$$c = (\ln(x) - m_X)s_Y^2 - r(\ln(y) - m_Y)s_Xs_Y$$

$$d = 2\pi xy\sqrt{s_X^2s_Y^2 - r^2s_X^2s_Y^2}$$

We have

$$r = \frac{\mu_X \mu_Y}{\sigma_X \sigma_Y} \left(-1 + e^{\rho \sqrt{\ln\left(1 + \frac{\sigma_X^2}{\mu_X^2}\right) \ln\left(1 + \frac{\sigma_Y^2}{\mu_Y^2}\right)}} \right) \sqrt{\ln\left(1 + \frac{\sigma_X^2}{\mu_X^2}\right) \ln\left(1 + \frac{\sigma_Y^2}{\mu_Y^2}\right)}$$

where μ_X and μ_Y are the means of X and Y , σ_X and σ_Y are their standard deviations and ρ their correlation coefficient.

Therefore,

$$f_L(x, y; \mu_X, \mu_Y, \sigma_X, \sigma_Y, \rho) = \frac{e^{\frac{ab+c}{d}}}{g}$$

where

$$\begin{aligned} a &= -2 \left(-1 + e^{\rho \sqrt{\ln\left(1 + \frac{\sigma_X^2}{\mu_X^2}\right) \ln\left(1 + \frac{\sigma_Y^2}{\mu_Y^2}\right)}} \right) \left(\ln(x) - \ln\left(\frac{\mu_X^2}{\sqrt{\mu_X^2 + \sigma_X^2}}\right) \right) \\ b &= m_X m_Y \sqrt{\ln\left(1 + \frac{\sigma_X^2}{\mu_X^2}\right) \ln\left(1 + \frac{\sigma_Y^2}{\mu_Y^2}\right)} \left(\ln(y) - \ln\left(\frac{\mu_Y^2}{\sqrt{\mu_Y^2 + \sigma_Y^2}}\right) \right) \\ c &= \left(\ln(y) - \ln\left(\frac{\mu_Y^2}{\sqrt{\mu_Y^2 + \sigma_Y^2}}\right) \right)^2 \sigma_X^2 + \left(\ln(x) - \ln\left(\frac{\mu_X^2}{\sqrt{\mu_X^2 + \sigma_X^2}}\right) \right)^2 \sigma_Y^2 \\ d &= 2 \left(\left(-1 + e^{\rho \sqrt{\ln\left(1 + \frac{\sigma_X^2}{\mu_X^2}\right) \ln\left(1 + \frac{\sigma_Y^2}{\mu_Y^2}\right)}} \right)^2 \ln\left(1 + \frac{\sigma_X^2}{\mu_X^2}\right) \ln\left(1 + \frac{\sigma_Y^2}{\mu_Y^2}\right) \mu_X^2 \mu_Y^2 - \sigma_X^2 \sigma_Y^2 \right) \\ g &= 2\pi xy \sqrt{- \left(-1 + e^{\rho \sqrt{\ln\left(1 + \frac{\sigma_X^2}{\mu_X^2}\right) \ln\left(1 + \frac{\sigma_Y^2}{\mu_Y^2}\right)}} \right)^2 \ln\left(1 + \frac{\sigma_X^2}{\mu_X^2}\right) \ln\left(1 + \frac{\sigma_Y^2}{\mu_Y^2}\right) \mu_X^2 \mu_Y^2 + \sigma_X^2 \sigma_Y^2} \end{aligned}$$

Appendix A.3.3. Gamma Distribution

(a) Univariate

The univariate Gamma distribution is a continuous probability distribution of a random variable X . The general form of its PDF is:

$$f_G(x; k, \theta) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-\frac{x}{\theta}}$$

where k is a shape parameter, θ a scale parameter and $\Gamma(u)$ the gamma function [52].

The mean μ and the standard deviation σ of X , are calculated as following:

$$\begin{aligned} \mu &= k\theta \\ \sigma &= k\theta^2 \end{aligned}$$

Therefore,

$$\begin{aligned} k &= \frac{\mu^2}{\sigma^2} \\ \theta &= \frac{\sigma^2}{\mu} \end{aligned}$$

and

$$f_G(x; \mu, \sigma) = \frac{1}{\Gamma\left(\frac{\mu^2}{\sigma^2}\right) \left(\frac{\sigma^2}{\mu}\right)^{\frac{\mu^2}{\sigma^2}}} x^{\left(\frac{\mu^2}{\sigma^2}-1\right)} e^{-\frac{x}{\mu^2}}$$

(b) Bivariate

The bivariate Gamma distribution is a continuous probability distribution of two variables X and Y. The copula version of its PDF is:

$$f_G(x, y; k_X, k_Y, \vartheta_X, \vartheta_Y, \rho) = \frac{ab}{c}$$

where

$$a = e^{(erfc^{-1}(2Q(k_Y, 0, \frac{y}{\vartheta_Y})))^2 + \frac{(-\rho erfc^{-1}(2Q(k_X, 0, \frac{x}{\vartheta_X})) + erfc^{-1}(2Q(k_Y, 0, \frac{y}{\vartheta_Y})))^2}{-1+\rho^2} - \frac{y}{\vartheta_Y} - \frac{x}{\vartheta_X}}$$

$$b = x^{-1+k_X} y^{-1+k_Y} \vartheta_Y^{-k_Y} \vartheta_X^{-k_X}$$

$$c = \sqrt{1-\rho^2} \Gamma(k_X) \Gamma(k_Y)$$

and k_X, k_Y are shape parameters, ϑ_X, ϑ_Y are scale parameters, and ρ the correlation coefficient of X and Y.

If μ_X and μ_Y are the means of X and Y, σ_X and σ_Y their standard deviations, and ρ their correlation coefficient, it can be shown that:

$$f_G(x, y; \mu_X, \mu_Y, \sigma_X, \sigma_Y, \rho) = \frac{ab}{c}$$

where

$$a = e^{(erfc^{-1}(2Q(\frac{\mu_Y^2}{\sigma_Y^2}, 0, \frac{y\mu_Y}{\sigma_Y^2})))^2 + \frac{(-\rho erfc^{-1}(2Q(\frac{\mu_X^2}{\sigma_X^2}, 0, \frac{x\mu_X}{\sigma_X^2})) + erfc^{-1}(2Q(\frac{\mu_Y^2}{\sigma_Y^2}, 0, \frac{y\mu_Y}{\sigma_Y^2})))^2}{-1+\rho^2} - \frac{x\mu_X}{\sigma_X^2} - \frac{y\mu_Y}{\sigma_Y^2}}$$

$$b = x^{(-1+\frac{\mu_X^2}{\sigma_X^2})} y^{(-1+\frac{\mu_Y^2}{\sigma_Y^2})} \sigma_X^{-\frac{2\mu_X^2}{\sigma_X^2}} \mu_X^{\frac{\mu_X^2}{\sigma_X^2}} \mu_Y^{\frac{\mu_Y^2}{\sigma_Y^2}} \sigma_Y^{-\frac{2\mu_Y^2}{\sigma_Y^2}}$$

$$c = \sqrt{1-\rho^2} \Gamma\left(\frac{\mu_X^2}{\sigma_X^2}\right) \Gamma\left(\frac{\mu_Y^2}{\sigma_Y^2}\right)$$

Appendix A.4. Copulas

If μ_X and μ_Y are the means of the variables X and Y, σ_X and σ_Y their standard deviations, and ρ their correlation coefficient, it can be shown that the bivariate PDFs of the other copulas of the program are defined as follows:

Appendix A.4.1. X: Normally Distributed—Y: Lognormally Distributed

$$f_{NL}(x, y; \mu_X, \mu_Y, \sigma_X, \sigma_Y, \rho) = \frac{e^c d}{g}$$

where

$$a = -\frac{2\ln(y) - 2\ln(\mu_Y) + \ln\left(1 + \frac{\sigma_Y^2}{\mu_Y^2}\right)^2}{2\sqrt{2\ln\left(1 + \frac{\sigma_Y^2}{\mu_Y^2}\right)}} - \frac{\left(2\ln(y) - 2\ln(\mu_Y) + \ln\left(1 + \frac{\sigma_Y^2}{\mu_Y^2}\right)\right)^2}{8\ln\left(1 + \frac{\sigma_Y^2}{\mu_Y^2}\right)}$$

$$b = - \frac{\left(\frac{2\ln(y) - 2\ln(\mu_Y) + \ln\left(1 + \frac{\sigma_Y^2}{\mu_Y^2}\right)}{2\sqrt{2}\sqrt{\ln\left(1 + \frac{\sigma_Y^2}{\mu_Y^2}\right)}} \sqrt{\ln\left(1 + \frac{\sigma_Y^2}{\mu_Y^2}\right)} \mu_Y + \rho\sigma_Y \operatorname{erfc}^{-1}\left(2Q\left(\frac{\mu_X^2}{\sigma_X^2}, 0, \frac{x\mu_X}{\sigma_X^2}\right)\right) \right)^2}{\ln\left(1 + \frac{\sigma_Y^2}{\mu_Y^2}\right) \mu_Y^2 - \rho^2\sigma_Y^2}$$

$$c = a + b - \frac{x\mu_X}{\sigma_X^2}$$

$$d = \left(\frac{x\mu_X}{\sigma_X^2}\right)^{\frac{\mu_X^2}{\sigma_X^2}}$$

$$g = xy\Gamma\left(\frac{\mu_X^2}{\sigma_X^2}\right) \sqrt{2\pi\ln\left(1 + \frac{\sigma_Y^2}{\mu_Y^2}\right) \left(1 - \frac{\rho^2\sigma_Y^2}{\ln\left(1 + \frac{\sigma_Y^2}{\mu_Y^2}\right) \mu_Y^2}\right)}$$

Appendix A.4.2. X: Lognormally Distributed—Y: Normally Distributed

$$f_{LN}(x, y; \mu_X, \mu_Y, \sigma_X, \sigma_Y, \rho) = \frac{e^c d}{g}$$

where

$$a = \operatorname{erfc}^{-1}\left(2Q\left(\frac{\mu_Y^2}{\sigma_Y^2}, 0, \frac{y\mu_Y}{\sigma_Y^2}\right)\right)^2 - \frac{\left(2\ln(x) - 2\ln(\mu_X) + \ln\left(1 + \frac{\sigma_X^2}{\mu_X^2}\right)\right)^2}{8\ln\left(1 + \frac{\sigma_X^2}{\mu_X^2}\right)}$$

$$b = - \frac{\left(\operatorname{erfc}^{-1}\left(2Q\left(\frac{\mu_Y^2}{\sigma_Y^2}, 0, \frac{y\mu_Y}{\sigma_Y^2}\right)\right) \sqrt{\ln\left(1 + \frac{\sigma_X^2}{\mu_X^2}\right)} \mu_X + \rho\sigma_X \left(\frac{2\ln(x) - 2\ln(\mu_X) + \ln\left(1 + \frac{\sigma_X^2}{\mu_X^2}\right)}{2\sqrt{2}\sqrt{\ln\left(1 + \frac{\sigma_X^2}{\mu_X^2}\right)}} \right) \right)^2}{\ln\left(1 + \frac{\sigma_X^2}{\mu_X^2}\right) \mu_X^2 - \rho^2\sigma_X^2}$$

$$c = a + b - \frac{y\mu_Y}{\sigma_Y^2}$$

$$d = \left(\frac{y\mu_Y}{\sigma_Y^2}\right)^{\frac{\mu_Y^2}{\sigma_Y^2}}$$

$$g = \left(\sqrt{2\pi}xy\Gamma\left(\frac{\mu_Y^2}{\sigma_Y^2}\right) \sqrt{\ln\left(1 + \frac{\sigma_X^2}{\mu_X^2}\right)} \sqrt{1 - \frac{\rho^2\sigma_X^2}{\ln\left(1 + \frac{\sigma_X^2}{\mu_X^2}\right) \mu_X^2}} \right)$$

Appendix A.4.3. X: Normally Distributed—Y: Gamma Distributed

$$f_{NG}(x, y; \mu_X, \mu_Y, \sigma_X, \sigma_Y, \rho) = \frac{e^{(erfc^{-1}(2Q(\frac{\mu_Y^2}{\sigma_Y^2}, 0, \frac{y\mu_Y}{\sigma_Y^2})) - \frac{(x-\mu_X)^2}{2\sigma_X^2} + \frac{(x\rho-\rho\mu_X+\sqrt{2}\sigma_Xerfc^{-1}(2Q(\frac{\mu_Y^2}{\sigma_Y^2}, 0, \frac{y\mu_Y}{\sigma_Y^2})))^2}{2(-1+\rho^2)\sigma_X^2} - \frac{y\mu_Y}{\sigma_Y^2}) \left(\frac{y\mu_Y}{\sigma_Y^2}\right)^{\frac{\mu_Y^2}{\sigma_Y^2}}}}{y\sigma_X\sqrt{2\pi(1-\rho^2)}\Gamma\left(\frac{y\mu_Y}{\sigma_Y^2}\right)}$$

Appendix A.4.4. X: Gamma Distributed– Y: Normally Distributed

$$f_{GN}(x, y; \mu_X, \mu_Y, \sigma_X, \sigma_Y, \rho) = \frac{e^{(-\frac{x\mu_X}{\sigma_X^2} + \frac{(y-\mu_Y+\sqrt{2}\rhoerfc^{-1}(2Q(\frac{\mu_X^2}{\sigma_X^2}, 0, \frac{x\mu_X}{\sigma_X^2}))\sigma_Y)^2}{2(-1+\rho^2)\sigma_Y^2}) \left(\frac{x\mu_X}{\sigma_X^2}\right)^{\frac{\mu_X^2}{\sigma_X^2}}}}{x\sigma_Y\sqrt{2\pi(1-\rho^2)}\Gamma\left(\frac{\mu_X^2}{\sigma_X^2}\right)}$$

Appendix A.4.5. X: Lognormally Distributed—Y: Gamma Distributed

$$f_{LG}(x, y; \mu_X, \mu_Y, \sigma_X, \sigma_Y, \rho) = \frac{e^c}{d}$$

where

$$a = \left(\frac{2\ln(y) - 2\ln(\mu_Y) + \ln\left(1 + \frac{\sigma_Y^2}{\mu_Y^2}\right)}{2\sqrt{2\ln\left(1 + \frac{\sigma_Y^2}{\mu_Y^2}\right)}} \right)^2 - \frac{\left(2\ln(y) - 2\ln(\mu_Y) + \ln\left(1 + \frac{\sigma_Y^2}{\mu_Y^2}\right)\right)^2}{8\ln\left(1 + \frac{\sigma_Y^2}{\mu_Y^2}\right)}$$

$$b = - \frac{\left(\left(-\ln(y) + \ln(\mu_Y) - \frac{\ln\left(1 + \frac{\sigma_Y^2}{\mu_Y^2}\right)}{2}\right)m_Y\sigma_X + \rho(x - \mu_X)\sigma_Y\right)^2}{2\sigma_X^2\left(\ln\left(1 + \frac{\sigma_Y^2}{\mu_Y^2}\right)\mu_Y^2 - \rho^2\sigma_Y^2\right)}$$

$$c = a + b - \frac{(x - \mu_X)^2}{2\sigma_X^2}$$

$$d = 2\pi y\sigma_X \sqrt{\ln\left(1 + \frac{\sigma_Y^2}{\mu_Y^2}\right) \left(1 - \frac{\rho^2\sigma_Y^2}{\ln\left(1 + \frac{\sigma_Y^2}{\mu_Y^2}\right)\mu_Y^2}\right)}$$

Appendix A.4.6. X: Gamma Distributed—Y: Lognormally Distributed

$$f_{GL}(x, y; \mu_X, \mu_Y, \sigma_X, \sigma_Y, \rho) = \frac{e^{a+b}}{c}$$

where

$$a = - \frac{\left(2\ln(x) - 2\ln(\mu_X) + \ln\left(1 + \frac{\sigma_X^2}{\mu_X^2}\right)\right)^2}{8\ln\left(1 + \frac{\sigma_X^2}{\mu_X^2}\right)}$$

$$b = - \frac{\left(\sqrt{\ln\left(1 + \frac{\sigma_X^2}{\mu_X^2}\right)}\mu_X(y - \mu_Y) - \rho\sigma_X\sigma_Y\left(\frac{2\ln(x) - 2\ln(\mu_X) + \ln\left(1 + \frac{\sigma_X^2}{\mu_X^2}\right)}{2\sqrt{\ln\left(1 + \frac{\sigma_X^2}{\mu_X^2}\right)}}\right)\right)^2}{2\sigma_Y^2\left(\ln\left(1 + \frac{\sigma_X^2}{\mu_X^2}\right)m_X^2 - \rho^2\sigma_X^2\right)}$$

$$c = 2\pi x\sigma_Y \sqrt{\ln\left(1 + \frac{\sigma_X^2}{\mu_X^2}\right)\left(1 - \frac{\rho^2\sigma_X^2}{\ln\left(1 + \frac{\sigma_X^2}{\mu_X^2}\right)m_X^2}\right)}$$

Appendix A.5. Nonparametric Distributions

Appendix A.5.1. Histograms

A histogram is a graphical representation of the distribution of a tuple of observed values of a variable X. If X is a continuous random variable, the histogram is an estimate of the probability distribution of X.

Appendix A.5.2. KDEs

Given a tuple of independent and identically distributed observed values (x_1, x_2, \dots, x_n) of a random variable X, the univariate KDE $\hat{f}_K(x; n, h)$ is defined as [53]:

$$\hat{f}_K(x; n, h) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right)$$

where:

1. n is the number of the observed values of the variable.
2. h is the bandwidth, a positive scalar that determines the width and smoothness of the kernel.
3. $K(u)$ is the kernel function.

Given two tuples of independent and identically distributed observed values (x_1, x_2, \dots, x_n) and (y_1, y_2, \dots, y_n) of two random variables X and Y, the bivariate KDE $\hat{f}(x, y; n, h_1, h_2)$ is defined as [53]:

$$\hat{f}(x, y; n, h_1, h_2) = \frac{1}{n|H|^{\frac{1}{2}}} \sum_{i=1}^n K\left((z - z_i)^T H^{-1}(z - z_i)\right)$$

where

$$z = \begin{bmatrix} x \\ y \end{bmatrix}$$

$$z_i = \begin{bmatrix} x_i \\ y_i \end{bmatrix}$$

$$H = \begin{bmatrix} h_1^2 & \rho h_1 h_2 \\ \rho h_1 h_2 & h_2^2 \end{bmatrix}$$

and ρ is the correlation coefficient of the two tuples.

The program uses the Gaussian kernel function:

$$K(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}$$

(a) Univariate KDE

$$\hat{f}(x; n, h) = \frac{1}{nh} \sum_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-x_i)^2}{2h^2}}$$

(b) Bivariate KDE

$$\hat{f}(x, y; n, h_1, h_2) = \frac{1}{2\pi n |H|^{\frac{1}{2}}} \sum_{i=1}^n e^{-\frac{1}{2}(z-z_i)^T H^{-1}(z-z_i)}$$

where

$$z = \begin{bmatrix} x \\ y \end{bmatrix}$$

$$z_i = \begin{bmatrix} x_i \\ y_i \end{bmatrix}$$

$$H = \begin{bmatrix} h_1^2 & \rho h_1 h_2 \\ \rho h_1 h_2 & h_2^2 \end{bmatrix}$$

Additional details concerning parametric and nonparametric distributions can be found in Supplementary File S2.

Appendix B

Appendix B.1. Software Availability and Requirements

B.1.1. Program name: Bayesian Diagnosis

B.1.2. Project home page: <https://www.hcsl.com/Tools/BayesianDiagnosis/> (accessed on 28 September 2023)

B.1.3. Operating systems: Microsoft Windows, Linux, Apple iOS

Programming language: Wolfram Language

B.1.4. Other software requirements:

1. For running the program: Wolfram Player[®], freely available at: <https://www.wolfram.com/player/> (accessed on 31 August 2023) or Wolfram Mathematica[®].
2. For editing the datasets: Wolfram Mathematica[®].

B.1.5. System requirements: Intel[®] i9[™] or equivalent CPU and 32 GB of RAM

B.1.6. License: Attribution—Noncommercial—ShareAlike 4.0 International Creative Commons License

Appendix C

Appendix C.1. A Note about the Program

Appendix C.1.1. About the Program Controls

The program features a tabbed user interface, designed to streamline user interaction and facilitate navigation across its multiple modules and submodules.

The numerical settings are defined by the user with sliders. Sliders can be finely manipulated by holding down the alt key or opt key while dragging the mouse. They can 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.

Appendix C.1.2. Range of input parameters

v : 0.010–0.500
 μ : 0.01–10,000.00
 σ : 0.01–3000.00
 ρ : −1.000–1.000
 h : 0.01–2.00
 x : 0.01–10,000.00
 y : 0.01–100,00.00

Appendix C.1.3. Datasets

The software is preloaded with the following datasets:

d1: Quantitative measurements of the first measurand (FPG) from diseased individuals (diabetic patients), aged 40–60.

d2: Quantitative measurements of the second measurand (HbA1c) from diseased individuals (diabetic patients), aged 40–60.

nd1: Quantitative measurements of the first measurand (FPG) from nondiseased individuals (nondiabetic patients), aged 40–60.

nd2: Quantitative measurements of the second measurand (HbA1c) from nondiseased individuals (nondiabetic patients), aged 40–60.

References

- Weiner, E.; Simpson, J.A.; Oxford University Press. *The Oxford English Dictionary*; Clarendon Press: Oxford, UK, 1989.
- Zweig, M.H.; Campbell, G. Receiver-Operating Characteristic (ROC) Plots: A Fundamental Evaluation Tool in Clinical Medicine. *Clin. Chem.* **1993**, *39*, 561–577. [[CrossRef](#)] [[PubMed](#)]
- Chatzimichail, T.; Hatjimihail, A.T. A Software Tool for Calculating the Uncertainty of Diagnostic Accuracy Measures. *Diagnostics* **2021**, *11*, 406. [[CrossRef](#)]
- Djulgovic, B.; van den Ende, J.; Hamm, R.M.; Mayrhofer, T.; Hozo, I.; Pauker, S.G.; International Threshold Working Group (ITWG). When Is Rational to Order a Diagnostic Test, or Prescribe Treatment: The Threshold Model as an Explanation of Practice Variation. *Eur. J. Clin. Investig.* **2015**, *45*, 485–493. [[CrossRef](#)] [[PubMed](#)]
- Choi, Y.-K.; Johnson, W.O.; Thurmond, M.C. Diagnosis Using Predictive Probabilities without Cut-Offs. *Stat. Med.* **2006**, *25*, 699–717. [[CrossRef](#)] [[PubMed](#)]
- Viana, M.A.G.; Ramakrishnan, V. Bayesian Estimates of Predictive Value and Related Parameters of a Diagnostic Test. *Can. J. Stat. Rev. Can. Stat.* **1992**, *20*, 311–321. [[CrossRef](#)]
- Gelman, A.; Carlin, J.B.; Stern, H.S.; Dunson, D.B.; Vehtari, A.; Rubin, D.B. *Bayesian Data Analysis*; CRC Press: Boca Raton, FL, USA, 2013.
- Van de Schoot, R.; Depaoli, S.; King, R.; Kramer, B.; Märtens, K.; Tadesse, M.G.; Vannucci, M.; Gelman, A.; Veen, D.; Willemsen, J.; et al. Bayesian Statistics and Modelling. *Nature Reviews Methods Primers* **2021**, *1*, 1–26. [[CrossRef](#)]
- Bours, M.J.L. Bayes' Rule in Diagnosis. *J. Clin. Epidemiol.* **2021**, *131*, 158–160. [[CrossRef](#)]
- Carlin, B.P.; Louis, T.A. *Bayesian Methods for Data Analysis*; CRC Press: Boca Raton, FL, USA, 2008.
- Martin, G.M.; Frazier, D.T.; Maneesoonthorn, W.; Loaiza-Maya, R.; Huber, F.; Koop, G.; Maheu, J.; Nibbering, D.; Panagiotelis, A. Bayesian Forecasting in Economics and Finance: A Modern Review. *Int. J. Forecast.* **2023**. [[CrossRef](#)]
- Liu, J.; Liu, S.J.; Wong, D.S.H. Process Fault Diagnosis Based on Bayesian Inference. In *Computer Aided Chemical Engineering*; Elsevier, Amsterdam, The Netherlands, Kraslawski, A., Turunen, I., Eds.; 2013; Volume 32, pp. 751–756.
- Dawid, A.P. Present Position and Potential Developments: Some Personal Views: Statistical Theory: The Prequential Approach. *J. R. Stat. Soc. Ser. A* **1984**, *147*, 278–292. [[CrossRef](#)]
- Lehmann, E.L.; Romano, J.P. *Testing Statistical Hypotheses*; Springer: New York, NY, USA, 2008.
- Box, G.E.P.; Cox, D.R. An Analysis of Transformations. *J. R. Stat. Society. Ser. B Stat. Methodol.* **1964**, *26*, 211–243. [[CrossRef](#)]
- D'Agostino, R.; Pearson, E.S. Tests for Departure from Normality. Empirical Results for the Distributions of b_2 and $\sqrt{b_1}$. *Biometrika* **1973**, *60*, 613–622. [[CrossRef](#)]
- Velanovich, V. Bayesian Analysis in the Diagnostic Process. *Am. J. Med. Qual. Off. J. Am. Coll. Med. Qual.* **1994**, *9*, 158–161. [[CrossRef](#)]
- Wilkes, E.H. A Practical Guide to Bayesian Statistics in Laboratory Medicine. *Clin. Chem.* **2022**, *68*, 893–905. [[CrossRef](#)]
- Geisser, S.; Johnson, W.O. *Modes of Parametric Statistical Inference*; John Wiley & Sons: Hoboken, NJ, USA, 2006.
- Spiegelhalter, D.J.; Abrams, K.R.; Myles, J.P. *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*; John Wiley & Sons Australia, Limited: Milton, QLD, Australia, 2004.
- Wilk, M.B.; Gnanadesikan, R. Probability Plotting Methods for the Analysis of Data. *Biometrika* **1968**, *55*, 1–17. [[CrossRef](#)] [[PubMed](#)]

22. ElSayed, N.A.; Aleppo, G.; Aroda, V.R.; Bannuru, R.R.; Brown, F.M.; Bruemmer, D.; Collins, B.S.; Gaglia, J.L.; Hilliard, M.E.; Isaacs, D.; et al. Classification and Diagnosis of Diabetes: Standards of Care in Diabetes-2023. *Diabetes Care* **2023**, *46* (Suppl. S1), S19–S40. [[CrossRef](#)] [[PubMed](#)]
23. Sun, H.; Saeedi, P.; Karuranga, S.; Pinkepank, M.; Ogurtsova, K.; Duncan, B.B.; Stein, C.; Basit, A.; Chan, J.C.N.; Mbanya, J.C.; et al. IDF Diabetes Atlas: Global, Regional and Country-Level Diabetes Prevalence Estimates for 2021 and Projections for 2045. *Diabetes Res. Clin. Pract.* **2022**, *183*, 109119. [[CrossRef](#)]
24. Centers for Disease Control and Prevention. National Center for Health Statistics, 2005–2016 National Health and Nutrition Examination Survey Data. 2005–2016. Available online: <https://wwwn.cdc.gov/nchs/nhanes/default.aspx> (accessed on 4 September 2023).
25. Centers for Disease Control and Prevention. National Health and Nutrition Examination Survey Questionnaire. 2005–2016. Available online: <https://wwwn.cdc.gov/nchs/nhanes/Search/variablelist.aspx?Component=Questionnaire> (accessed on 4 September 2023).
26. Menke, A.; Rust, K.F.; Savage, P.J.; Cowie, C.C. Hemoglobin A1c, Fasting Plasma Glucose, and 2-Hour Plasma Glucose Distributions in U.S. Population Subgroups: NHANES 2005–2010. *Ann. Epidemiol.* **2014**, *24*, 83–89. [[CrossRef](#)]
27. Silverman, B.W. *Density Estimation for Statistics and Data Analysis*; CRC Press: Boca Raton, FL, USA, 1986.
28. Obermeyer, Z.; Emanuel, E.J. Predicting the Future—Big Data, Machine Learning, and Clinical Medicine. *N. Engl. J. Med.* **2016**, *375*, 1216–1219. [[CrossRef](#)] [[PubMed](#)]
29. Topol, E.J. Individualized Medicine from Prewomb to Tomb. *Cell* **2014**, *157*, 241–253. [[CrossRef](#)]
30. Tucker, L.A. Limited Agreement between Classifications of Diabetes and Prediabetes Resulting from the OGTT, Hemoglobin A1c, and Fasting Glucose Tests in 7412 U.S. Adults. *J. Clin. Med. Res.* **2020**, *9*, 2207. [[CrossRef](#)]
31. Smith, A.F.M.; Gelfand, A.E. Bayesian Statistics without Tears: A Sampling-Resampling Perspective. *Am. Stat.* **1992**, *46*, 84–88.
32. O’Hagan, A.; Buck, C.E.; Daneshkhah, A.; Eiser, J.R.; Garthwaite, P.H.; Jenkinson, D.J.; Oakley, J.E.; Rakow, T. *Uncertain Judgements: Eliciting Experts’ Probabilities*; John Wiley & Sons: Hoboken, NJ, USA, 2006.
33. Berger, J.O. *Statistical Decision Theory and Bayesian Analysis*; Springer Science & Business Media: New York, NY, USA, 1985.
34. McGrayne, S.B. *The Theory That Would Not Die: How Bayes’ Rule Cracked the Enigma Code, Hunted Down Russian Submarines, & Emerged Triumphant from Two Centuries of C*; Yale University Press: New Haven, CT, USA, 2011.
35. Box, G.E.P.; Tiao, G.C. *Bayesian Inference in Statistical Analysis*; John Wiley & Sons, Incorporated: Hoboken, NJ, USA, 2011.
36. Tamrakar, S.; Choubey, S.B.; Choubey, A. *Computational Intelligence in Medical Decision Making and Diagnosis: Techniques and Applications*; CRC Press: Boca Raton, FL, USA, 2023.
37. Wasserman, L. *All of Nonparametric Statistics*; Springer Science & Business Media: New York, NY, USA, 2006.
38. Pearl, J.A. Probabilistic Calculus of Actions. In *Uncertainty Proceedings 1994*; Lopez de Mantaras, R., Poole, D., Eds.; Morgan Kaufmann: San Francisco, CA, USA, 1994; pp. 454–462.
39. Heckerman, D.; Geiger, D.; Chickering, D.M. Learning Bayesian Networks: The Combination of Knowledge and Statistical Data. *Mach. Learn.* **1995**, *20*, 197–243. [[CrossRef](#)]
40. American Diabetes Association. Classification and Diagnosis of Diabetes: Standards of Medical Care in Diabetes—2021. *Diabetes Care* **2021**, *44* (Suppl. S1), S15–S33. [[CrossRef](#)]
41. Meneilly, G.S.; Elliott, T. Metabolic Alterations in Middle-Aged and Elderly Obese Patients with Type 2 Diabetes. *Diabetes Care* **1999**, *22*, 112–118. [[CrossRef](#)] [[PubMed](#)]
42. Geer, E.B.; Shen, W. Gender Differences in Insulin Resistance, Body Composition, and Energy Balance. *Gen. Med.* **2009**, *6* (Suppl. S1), 60–75. [[CrossRef](#)] [[PubMed](#)]
43. Van Cauter, E.; Polonsky, K.S.; Scheen, A.J. Roles of Circadian Rhythmicity and Sleep in Human Glucose Regulation. *Endocr. Rev.* **1997**, *18*, 716–738.
44. Colberg, S.R.; Sigal, R.J.; Fernhall, B.; Regensteiner, J.G.; Blissmer, B.J.; Rubin, R.R.; Chasan-Taber, L.; Albright, A.L.; Braun, B. Exercise and Type 2 Diabetes: The American College of Sports Medicine and the American Diabetes Association: Joint Position Statement. *Diabetes Care* **2010**, *33*, e147–e167. [[CrossRef](#)]
45. Salmerón, J.; Manson, J.E.; Stampfer, M.J.; Colditz, G.A.; Wing, A.L.; Willett, W.C. Dietary Fiber, Glycemic Load, and Risk of Non-Insulin-Dependent Diabetes Mellitus in Women. *JAMA J. Am. Med. Assoc.* **1997**, *277*, 472–477. [[CrossRef](#)]
46. Surwit, R.S.; van Tilburg, M.A.L.; Zucker, N.; McCaskill, C.C.; Parekh, P.; Feinglos, M.N.; Edwards, C.L.; Williams, P.; Lane, J.D. Stress Management Improves Long-Term Glycemic Control in Type 2 Diabetes. *Diabetes Care* **2002**, *25*, 30–34. [[CrossRef](#)]
47. Pandit, M.K.; Burke, J.; Gustafson, A.B.; Minocha, A.; Peiris, A.N. Drug-Induced Disorders of Glucose Tolerance. *Ann. Intern. Med.* **1993**, *118*, 529–539. [[CrossRef](#)]
48. Dupuis, J.; Langenberg, C.; Prokopenko, I.; Saxena, R.; Soranzo, N.; Jackson, A.U.; Wheeler, E.; Glazer, N.L.; Bouatia-Naji, N.; Gloyn, A.L.; et al. New Genetic Loci Implicated in Fasting Glucose Homeostasis and Their Impact on Type 2 Diabetes Risk. *Nat. Genet.* **2010**, *42*, 105–116. [[CrossRef](#)]
49. Haekel, R.; Wosniok, W.; Arzideh, F. A Plea for Intra-Laboratory Reference Limits. Part 1. General Considerations and Concepts for Determination. *Clin. Chem. Lab. Med. CCLM/FESCC* **2007**, *45*, 1033–1042. [[CrossRef](#)] [[PubMed](#)]

50. Arzideh, F.; Wosniok, W.; Gurr, E.; Hinsch, W.; Schumann, G.; Weinstock, N.; Haeckel, R. A Plea for Intra-Laboratory Reference Limits. Part 2. A Bimodal Retrospective Concept for Determining Reference Limits from Intra-Laboratory Databases Demonstrated by Catalytic Activity Concentrations of Enzymes. *Clin. Chem. Lab. Med. CCLM/FESCC* **2007**, *45*, 1043–1057. [[CrossRef](#)] [[PubMed](#)]
51. Centers for Disease Control and Prevention. National Center for Health Statistics NHANES—NCHS Research Ethics Review Board Approval. 2023. Available online: <https://www.cdc.gov/nchs/nhanes/irba98.htm> (accessed on 4 September 2023).
52. Forbes, C.; Evans, M.; Hastings, N.; Peacock, B. *Statistical Distributions*; John Wiley & Sons: Hoboken, NJ, USA, 2011.
53. Gramacki, A. *Nonparametric Kernel Density Estimation and Its Computational Aspects*; Springer: Berlin/Heidelberg, Germany, 2017.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.