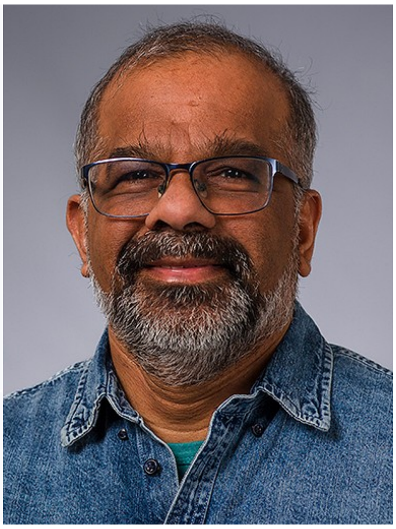
Σ Statlytika

Statistical and Data Analytics



https://priyantha-wijayatunga.github.io/statlytika/


**Priyantha Wijayatunga, PhD**

**Contact:**

*Address: Triangelgatan 11, 90752 Umeå, Sweden*
*Tel: 0046 703254262*
*Email: priyantha.wijayatunga@gmail.com*
*LinkedIn: https://www.linkedin.com/in/priyantha-wijayatunga-09b8b939/*
*SciProfiles: https://sciprofiles.com/profile/priyantha-wijayatunga*
*ORCID: https://orcid.org/0000-0003-1654-9148*


# Σ Statlytika

## *Transforming Data into Evidence-Based Insights for Decisive Actions, Sustainable Growth and Excellence*

## Our Services:

### Statistical Consultancy for Growth and Excellence

*We Transform Your Data into Strategic Advantage!*

When you possess valuable datasets and information, professional statistical consultancy empowers you to utilize them in productive and effective ways. By extracting meaningful knowledge and evidence from your data, we help improve your business outcomes and enhance your scientific, socio-economic, and policy project work.

Our technical statistical expertise does exactly this. We cater to your needs by elevating your scientific research, business operations, socio-economic studies, policy development, and other projects through the discovery of actionable insights within your datasets.

The knowledge we extract creates intelligence and wisdom in your field of work, paving the way for sustainable growth and helping you thrive in competitive environments. Our team analyzes your data using modern statistical methods, advanced data analytical techniques, and cutting-edge machine learning tools.

Through rigorous analysis, we discover:

- New patterns and trends
- Significant associations
- Causal relationships
- Underlying regularities

These insights enable you to perform targeted interventions, make accurate predictions, develop robust explanations, and draw reliable conclusions that drive your objectives forward.

Partner with us and get our service to transform your raw data and information into strategic advantage and evidence-based decision-making power.

## Educational Services

*We Elevate Your Analytics Capabilities!*

Are you looking to enhance your statistical and data analytical skills to better manage your business, accelerate scientific and industrial growth, or develop impactful socio-economic projects? We offer tailored educational experiences designed to help you thrive in your field.

Our specialized programs include:

- Custom courses and workshops
- Interactive discussion sessions
- Targeted seminars and masterclasses

Curriculum Highlights:

- Exploratory data analysis techniques
- Advanced regression modeling
- Predictive analytics and forecasting
- Clustering and classification methodologies
- Comprehensive risk analysis
- Uncertainty reasoning and probabilistic modeling

We employ cutting-edge statistical methods, data science frameworks, and machine learning tools throughout our educational offerings. Each program is customized to align with your specific needs and objectives.

Choose from flexible delivery options:

- In-person immersive sessions
- Interactive online learning
- Hybrid approaches

This flexibility allows you to select the learning pace and schedule that works best for your team. Contact us today to design a customized educational program that addresses your unique analytical challenges.

## Research Support

*We Provide Expert Guidance for Academic Excellence!*

We specialize in providing exceptional support to undergraduate and graduate students with their research projects, theses, and academic essays. Our expertise enables students to elevate their analytical capabilities and produce superior academic work.

Our research support services include:

- •Instruction in advanced statistical methodologies relevant to specific research questions
- •Custom software coding tutorials tailored to unique analysis requirements
- •Step-by-step demonstrations of analytical techniques
- •Guidance on interpreting and presenting research findings effectively

Our approach emphasizes both theoretical understanding and practical application, ensuring students develop transferable skills that extend beyond their immediate academic projects. We pride ourselves on helping students develop their independent analytical, logical and critical thinking skills while meeting the rigorous standards of academic research.

# Priyantha Wijayatunga, PhD

## Educational Background

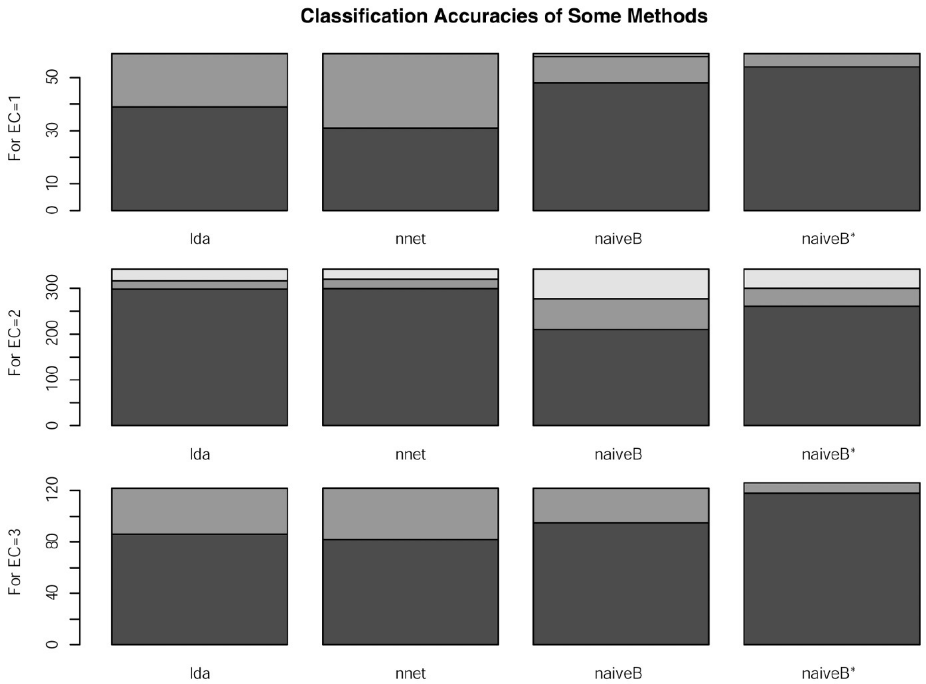
- Ph.D. Degree in Mathematical and Computing Sciences (Tokyo Institute of Technology Tokyo Japan, 2007)
- M.Sc. Degree in the same subject from the same Institute.
- Diploma in Mathematical Statistics (Cambridge University, England, 1998)
- B.Sc. (Special) Degree in Mathematics (University of Kelaniya, Sri Lanka, 1993)

## Work Experience

- **Associate Professor** in Statistics (Umeå University, Sweden, 2008.04.01-2024.12.31)
- **Postdoctoral Fellow** in Bioinformatics (Tokyo Institute of Technology, 2007.04.01-2008.03.31)
- **Lecturer** (Wayamba University, Sri Lanka, 1996.08.01-2001.03.31)
- **Actuarial Assistant** (Union Assurance Ltd, Sri Lanka, 1994.08.01-1996.07.31)
- **Assistant Lecturer** (University of Kelaniya, Sri Lanka, 1993.08.01-1994.07.31)

## Our Scientific Expertise

**Efficient prediction of companies at risk:** I have shown that when predictions are done where predicting some events that are minority are more important than the others that are the majority, probabilistic prediction models such as naive Bayesian networks can be constructed in such a way that we get higher prediction accuracies for important events while losing a little bit of prediction accuracy for unimportant events. My novel model (naiveB*) attained 92% accuracy of prediction for companies that are at risk of bankruptcy (EC=1) and 93% of accuracy of prediction for those are not at risk of bankruptcy at all (EC=3). All other models such as neural networks fail to attain such high prediction accuracies for the important events, but work well for unimportant events, which is not desired for investors and other business professionals. This model can be used to predict rare and important events of interest in many business, scientific, etc. contexts.




Dark areas shows the correct predictions. Notice the right column that is my new model

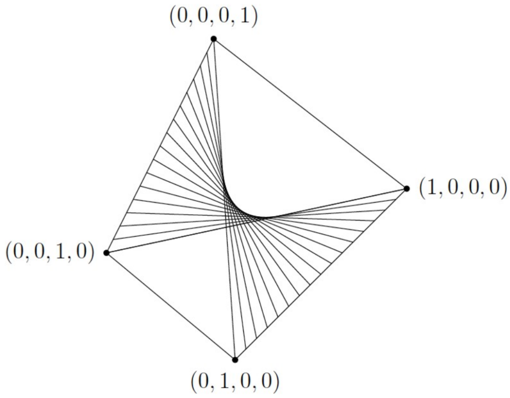
**Predicting severe medical conditions of patients with traumatic brain injury:** I have developed an effective and operational method to predict accurately, rare or extreme events happening over time using multiple time series data. When it is needed to predict relatively rare events of elevated intracranial blood pressure (ICP) of patients with brain traumatic injuries within the next hour or so from their clinical condition of the last few hours, I developed a probabilistic model to do it with highest accuracy. The idea is not just to use raw data, but extract features from them, e.g., oscillation patterns, spikes, etc. in the selected time intervals of the past. Severe elevations of ICP, that are rare are difficult to predict with traditional prediction models such as neural networks, etc. So, I invented a novel method which calibrate and weigh predictive probabilities in order to select predictions. The method works in such a way that it gives enhanced predictions for the rare events of severe levels of ICP (the ICP level 2) while losing slightly prediction accuracy for normal levels of ICP (the ICP level 1), for the first to the sixth ten-minute time intervals into the future. The weighing of predictive probabilities is done in order to select predictions through minimizing a prediction error loss function that has different losses for different errors.

| | | Specificity (%), true ICP = "1" | | Sensitivity (%), true ICP = "2" | |
|---|---|---|---|---|---|
| | **10-min interval** | **General[a]** | **Enhanced[b]** | **General[a]** | **Enhanced[b]** |
| Probability of predicting the true ICP level (%) | First | 96.0 (95.8; 96.2) | 95.0 (94.9; 95.1) | 85.4 (83.6; 85.1) | 87.1 (86.3; 87.7) |
| | Second | 96.1 (95.9; 96.3) | 94.7 (94.5; 94.9) | 74.7 (73.5; 75.8) | 80.6 (80.1; 81.2) |
| | Third | 95.6 (95.5; 95.7) | 93.8 (93.6; 93.9) | 72.1 (69.8; 75.6) | 75.7 (74.8; 76.7) |
| | Fourth | 95.5 (95.3; 95.7) | 93.0 (92.9; 93.2) | 67.4 (66.3; 68.4) | 75.0 (74.3; 75.9) |
| | Fifth | 95.2 (95.0; 95.5) | 93.3 (93.0; 93.4) | 67.3 (66.0; 68.6) | 73.0 (72.4; 73.6) |
| | Sixth | 95.1 (94.9; 95.2) | 93.9 (93.6; 94.1) | 67.0 (66.3; 67.8) | 72.7 (72.1; 73.4) |

The predictive accuracies for ICP level 2 are enhanced from general predictions

The model is probabilistic therefore it can incorporate other information types as past ECG signal variation, respiration pattern, blood pressure variation, etc. of the patient. In fact, in this work no such extra knowledge could improve prediction accuracies. But my proposed method can be useful in other cases. This model can be used to predict stock market, exchange rate, interest rate, etc. prediction from multiple sources of data. And it is applicable for prediction of medical conditions of patients with clinical information over the time. And so on.

**Building a non-linear correlation measure:** Using a general form of probability distance metric, namely, Hellinger distance I obtained a generalized version of the Pearson's correlation coefficient for two ordinal variables such as educational status and socio-economic status of people or similar cases. It can measure any type of stochastic dependence accurately. For continuous variables like age, weight, etc. optimal categorization is needed, as in case of popular maximal information coefficient (MIC). However the measure I obtained is more accurate. And I have shown that any good measure of dependence is computational, even if it has a closed form formula. High level of accuracy is obtained through, on one hand, using an accurate normalization for probability distance used to build the measure, that is the distance between the existing phenomenon of dependence and its assumed independence. And on the other hand, using a general metric distance that is independent of the nature of the dependence. Some researchers have implemented it in R and made available over the internet: wcor: Wijayatunga coefficient, https://rdrr.io/github/vthorrf/wijayatunga/man/wcor.html. The measure can be useful in genetic association studies, socio-economic studies where sample survey data consist of a lot of categorical variables.




Geometrical figure of joint probability distribution of binary X and Y. Surface of lines inside shows the independence and vertices are the deterministic dependencies of them

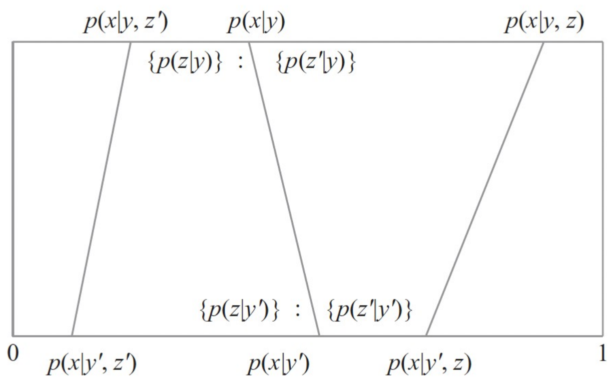
**Defining outcome scores for causal inference:** Balancing scores such as prognostic scores are popular in causal inference tasks to remove confounding bias in the given causal relationships. Using a probabilistic analysis, I have shown how to derive prognostic scores correctly, which I call outcome scores. They are more accurate than currently popular prognostic scores. Also, I have shown how to estimated causal effect accurately by combining the prognostic and the propensity scores. This gives more accurate causal effect estimates.

**Showing connection between causal effect estimators:** I have shown the equivalence between the two popular frameworks of causal inference for observational data, namely, causal graphical model and potential outcome causal model. This had been partially discussed previously and my analysis gave a complete picture that is useful for empirical researchers. I have shown how to obtain many of the popular causal effect estimators found in the applications of the potential outcome causal model from the application of the causal graphical models. This is important as many researchers believe that the two frameworks are independent. Also I showed how to derive doubly robust estimator mathematically. It is claimed by the inventors of this estimator that it is not easy to derive it. My mathematical derivation is useful for understanding the properties of the estimator.

**Selection of confounders and handing M-bias:** Causal effect estimation with observational data needs to consider a sufficient subset of confounders of the causal relation of interest by controlling for (conditioning on) them. The current practice of selecting the subset is according to their predictive ability of the treatment firstly, and then the outcome. But I showed that it should be done other way round; firstly they should be predictive of the outcome and then the treatment. This is important in order to obtain unbiased estimates of causal effects efficiently and accurately. Furthermore I showed how to handle associative confounders (those are not causally affecting both the treatment and outcome, but associated with them). Currently there is no clear consensus about handling them. Note that associative confounders are said to cause so-called M-bias. I showed that it is beneficial to condition on associative confounders when they are strongly dependent with both the treatment and outcome whereas it is not so when they are weakly dependent. In either case, there is confounding bias but through my proposal it can be minimized. These results are important empirical causal effect estimation tasks.

**Resolving Controversy between Significance Testing and Bayesian Testing (Jeffreys-Lindley's paradox):** Mainly there are two hypothesis testing frameworks, namely, Fisherian significance testing and Bayesian hypothesis testing. Jeffreys-Lindley paradox is a case where they contradict with each other. This has caused confusion among data analysts for selecting a methodology for their statistical inference tasks. Though the paradox goes back to mid 1930's so far there hasn't been a satisfactory resolution given for it. I have showed that it arises mainly due to the simple fact that, in the significance testing the difference between the hypothesized parameter value and the observed estimate of the parameter is assessed in terms of the standard error of the estimate, no matter what the actual numerical difference is and how small the standard error is, whereas in the Bayesian methodology it has no effect due to the definition of the Bayes factor in the context, even though such an assessment is present. In fact, the paradox is an instance of conflict between statistical and practical significance and a result of using a sharp null hypothesis to approximate an acceptable small range of values for the parameter.

**Resolving Simpson's paradox:** Simpson's paradox is the case where association between a factor and an outcome negates when another factor is considered. I have shown a simple diagram that is useful for understanding this phenomenon. And the diagram shows that this process is never ending. Also it shows that conditional associations between discrete variables are uncertain—they depend on what extra factors included, i.e., they can alternate from positive to negative and vice versa. It can be shown that it is not the case for continuous outcomes—there is a limit of such negation, at least practically. These findings are important when doing predictions.




An occurrence of the Simpson's paradox: Negative association between binary $X$ and $Y$ is negated to two positive associations when another binary variable $Z$ is considered. {a}:{1-a} the ratio of lengths of respective line segments.