
A relevant Weakness in Survival Machine Learning Models: non-Ergodicity.

WORKING PAPER

Diego Vallarino^{1*}

¹Independent Researcher, Madrid, Spain.

*Corresponding author

diego.vallarino@gmail.com

June 2023

Abstract:

This study examines ergodicity and survival analysis. The ergodic theorem concerns statistical phenomena's generalizability. Ergodicity detects statistical incongruity and inference mistakes like Simpson's dilemma and the ecological fallacy, according to the research. The article discusses ergodicity's idea, consequences for difficult datasets, and potential to improve survival rate estimations and forecasts. Survival analyses and data interpretation improve with non-ergodicity limitations. We have demonstrated that ergodicity analysis is straightforward to assess if the data collection is appropriately planned during experiment design to provide us a longitudinal and continuous view of each agent (person, corporation, machine, etc.).

Keywords: Ergodicity, Survival analysis, Statistical inference, Survival Machine Learning, Group-to-individual generalizability.

MSC2020: 62P10, 92B15

1. Introduction

The ergodic theorem is a broad and formal mathematical formulation that deals with statistical phenomena' generalizability across levels and units of study. According to ergodic theory, the patterns of interindividual and intraindividual variation in human subjects' data are asymptotically comparable, which is a required but not necessarily sufficient condition for ergodicity (see [1]).

The ergodic theorem may be seen as a broad framework for identifying particular situations of statistical incongruity and inference mistakes, such as Simpson's dilemma and the ecological fallacy. Simpson's paradox (see [2]) is a statistical phenomenon in which subgroup trends diverge from (or are even inverse to) the aggregate trend when the groups are merged. The ecological fallacy is a frequent and troublesome statistical interpretation mistake that occurs when statistical results from groups are improperly extrapolated to individuals (see [3]).

Hamaker ([4] [5]), who discusses the relationship between typing speed and errors, provides a simple example. The connection is negative at the group level, since experienced typists are both quicker and more competent. Within individuals, however, the connection is positive. The quicker a person types, the more errors she or he makes in comparison to their own proficiency at slower rates. As a result of the data aggregation, we get an example of Simpson's dilemma, and we commit an ecological fallacy if we conclude that the association seen at the group level reflects any of the people in the group. Simpson's paradox and the ecological fallacy both remind us that the individual and group levels are not always connected. Before making any extrapolations, the consequences of nonergodicity in a specific dataset should be explicitly examined.

This paper will explore the concept of ergodicity and its importance in survival analysis. In this way, the second section of the paper will delve into the conceptualization of ergodicity. It will discuss the two strict conditions required for the generalization of observations across individuals: population homogeneity and stationarity. The violation of these conditions will be illustrated through examples from personality tests and emotional experience development studies. These violations demonstrate that inter-individual variation cannot be equated with intra-individual variation due to non-ergodicity. The section will emphasize the importance of understanding ergodicity in analyzing complex systems and processes accurately.

The third section will focus on how ergodicity analysis can enhance survival analysis. It will explore how considering ergodicity allows researchers to estimate long-term survival rates more accurately, especially in cases where survival rates are not constant over time or when competing risks are involved. In the fourth part of the paper, we approach to mention how ergodicity affects the machine learning models that are used for survival analysis.

This paper will emphasize the significance of considering ergodicity in survival analysis and machine learning models. It will provide insights into the conceptualization of ergodicity, its implications for analyzing complex datasets, and how ergodicity analysis can improve the accuracy of survival rate estimations and predictions. By addressing the limitations caused by non-ergodicity, researchers can enhance the validity and reliability of survival analyses and draw more meaningful conclusions from their data.

2. An approximation to the ergodicity.

Unfortunately, applied ergodicity tests are rare in the social, behavioral, and medical sciences. While others have noted that processes inside individuals vary from processes sampled across individuals over time ([6] [7] [8]), assessing the magnitude and possible impact of this mismatch in psychological and medical domains should be a regular focus of scientific investigation. While Pearl ([2]) proved that there is no single diagnosis or remedy for Simpson's dilemma, we suggest a reasonably simple method for directly testing for nonergodicity and, hence, group-to-individual generalizability in statistical studies.

Simply put, comparisons of the first and second moments (mean and variance) of intraindividual and interindividual distributions may provide information on the correctness of group and individual generalizations. Prodigious collaboration efforts across all domains of human subjects research would be necessary to properly study group-to-individual generalizability throughout the social and medical sciences. Individual researchers may address the appropriateness of their data for generalizations from aggregated findings to individual participants in the meanwhile by using appropriate study methodologies and data gathering paradigms.

Scientists who want to generalize results across interindividual and intraindividual levels of analysis, in particular, should gather several measurements inside participants throughout time—whether or not the study objective is explicitly longitudinal. Furthermore, sharing data and findings might reduce the burden of testing for ergodicity in future investigations. Fortunately, as data resources become more widely accessible via open access, we can begin to address this issue collaboratively. To assess the significance of this endeavor, we compare intraindividual and interindividual variance in six separate datasets of frequently sampled people.

One of the difficulties related to handling complex datasets is the need to consider the ergodicity in the training samples ([8] [9]). A system is ergodic if its expectation value (the average of many independent systems running the experiment) is equal to its long-run average (the average of a single system running the experiment repeatedly, maintaining its state from one sample to the next), so that their average statistical properties can be deduced from a single large enough random sample of the system's behaviour ([9] [10] [11]).

The importance of ergodicity lies in the scope of the conclusions that we can draw from the analysis. When we are dealing with non-ergodic sets, the characteristics of the set cannot be used to infer something about a specific individual from that set.

In the case of survival analysis, as reflected in the literature analysed, conclusions are drawn at the group level, using averages, and at the individual level, without previously analysing the ergodicity of the training dataset (e.g. [12] [8] [4] [5] [6] [13]). This presents an important weakness at the level of survival analysis, since nothing can be affirmed at the individual level, if it is not first confirmed that the machine learning models comply with the classical ergodic theorem.

What is normally done in survival analyses using machine learning models, with large volumes of data, is to segment the population and ensure that all segments are represented. Data are then obtained from a small sample that is assumed to be representative. Since it is not an ergodic set, the results will not coincide.

This is what we usually know as the margin of error in the analysis. This error is not an error in the literal sense but refers to the expected difference due to the non-ergodicity of the set. Although many times (not to say most cases) the margin of error represents more a misuse of statistical concepts than a probability of "non-adjustment" of the inference.

For example, if a training data set is not representative of an ergodic process, a model trained on this data set may not accurately predict future outcomes or may have poor generalization performance when applied to new data. A similar problem is the one that scientists in general have when they try to infer general laws from concrete experiments. When is it correct to generalise and when is it not? The answer depends on ergodicity.

3. Ergodicity Conceptualization in general models

Molenaar and Campbell [6] argued that the classical ergodic theorem requires that the generalization of observations across individuals can only be done under two strict conditions.

The first condition is that the population must be homogeneous and the same statistical model that is used to describe the group as a whole must be applied to all subjects in the population. In other words, the means and other descriptive statistics that describe the data should not vary between individual participants. Only then can the statistical model of the population be applied to an individual participant in that population.

To illustrate violations of ergodicity, Molenaar and Campbell ([6]) referred to a repeated measurement of a personality test that 22 participants completed for 90 consecutive days. The questionnaire consisted of 30 items to assess the factors that represent the components of the Big Five personality factors (Neuroticism, Extraversion, Agreeableness, Conscientiousness, and Intellect). Group analysis showed that the questionnaire reliably explained the Big Five personality component factors. However, when looking at the 30 repeatedly measured item scores of each of the individual participants, the Big Five personality factors do not reliably explain the correlations between the scores. The factor loadings were substantially different for each of the individual test participants, both in terms of the number of factors involved and how the factors were related to the questionnaire items.

The second condition for ergodicity is stationarity. It requires that the data be stable and that the mean and variance do not change between measurements. In other words, statistical parameters such as factor loadings must remain the same across all measurements over time. Molenaar, Sinclair, Rovine, Ram, & Corneal ([14]) argued that virtually all studies that focus on change over time in psychological characteristics within individuals violate the stationarity condition for the ergodicity of the data. They stated that the combination of individuals in groups is inappropriate for developmental studies since developmental processes are almost always non-stationary and therefore non-ergodic.

They illustrated this point with data from a study that investigated the emotional experience development of eight children and eight stepchildren as they interacted with their parents for 80 interactions over time. For each participant, a factor analysis was used to identify three factors: Involvement, Anger, and Anxiety. The authors fitted an unsteady state-space model to single-subject time series data using a recursive estimator (EFKIS).

The time series model showed that the relationship between anxiety and involvement was dynamic, changing from a negative to a positive relationship about halfway through the

time series. Their study clearly showed that due to the violation of this ergodicity condition, inter-individual variation cannot be equated with an intra-individual variation.

Ergodicity is a property of a system that describes how its statistical properties change over time. In an ergodic system, the long-term statistical properties of the system can be inferred from a single, long-run observation of the system. In other words, if you observe the system for a long enough time, you can determine its statistical properties with a high degree of accuracy.

On the other hand, if a system is non-ergodic, the long-term statistical properties of the system cannot be inferred from a single, long-run observation of the system. This means that it is not possible to determine the statistical properties of the system by observing it for a long time, and it is necessary to make multiple observations of the system to accurately determine its statistical properties.

3.1 Our analysis

We generated a simulated dataset for studying the survival or time-to-event outcomes in a group of individuals. Here is a detailed explanation of the code:

n <- 100: This line assigns the value 100 to the variable **n**, representing the number of individuals in the dataset. Each individual will have survival-related information.

time <- 1:90: This line creates a sequence from 1 to 90 and assigns it to the variable **time**. It represents the time points at which survival-related events or observations are recorded for each individual.

This **time** variable, or data point, is critical to be able to measure the existence of ergodicity or not. If it is not created, the possibility of calculating it directly is lost.

anxiety <- matrix(rnorm(n * length(time), mean = 0, sd = 1), nrow = n): This code generates a matrix called **anxiety**. It contains random numbers drawn from a normal distribution with a mean of 0 and a standard deviation of 1. The matrix has **n** rows (representing individuals) and **length(time)** columns (representing time points). These random numbers represent anxiety coefficients for each individual at each time point.

The subsequent lines generate additional health-related variables that may potentially impact survival outcomes. These variables are also generated using random numbers drawn from normal, chi-square, poisson, exponential, and logistic distributions, with specific mean and standard deviation values assigned to each variable. The variables include **stress_levels**, **genetic_predisposition**, **past_traumatic_experiences**, **socioeconomic_status**, **social_support_network**, **coping_mechanisms**, **personality_traits**, **environmental_factors**, **health_conditions**, and **life_events**. Each variable has a corresponding mean and standard deviation, which can be interpreted in the context of survival analysis.

Finally, the dataset is created by combining all the generated variables into a data frame named **data**. The **data.frame()** function is used to create the data frame, with each variable assigned as a column. Additionally, the **time** variable is repeated **n** times to match the number of rows in the dataset. This allows for associating the respective survival times or observations with each individual and their corresponding health-related variables.

Below are two graphs illustrating the distribution of two variables, Anxiety and Stress, and how they act at each time (90 moments) for each individual (100 in total).

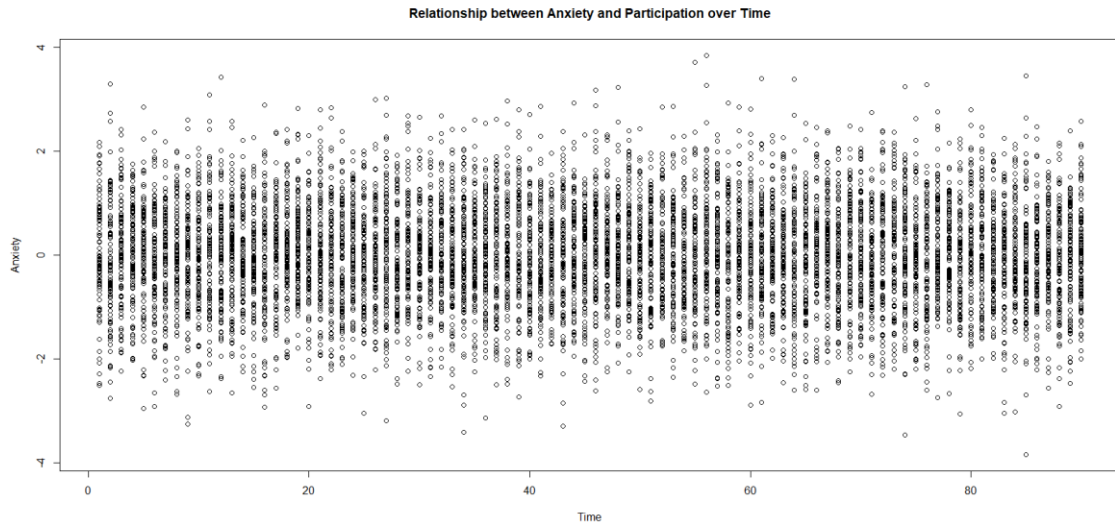


Figure 1: Plot the relationship between anxiety and participation over time for each participant.

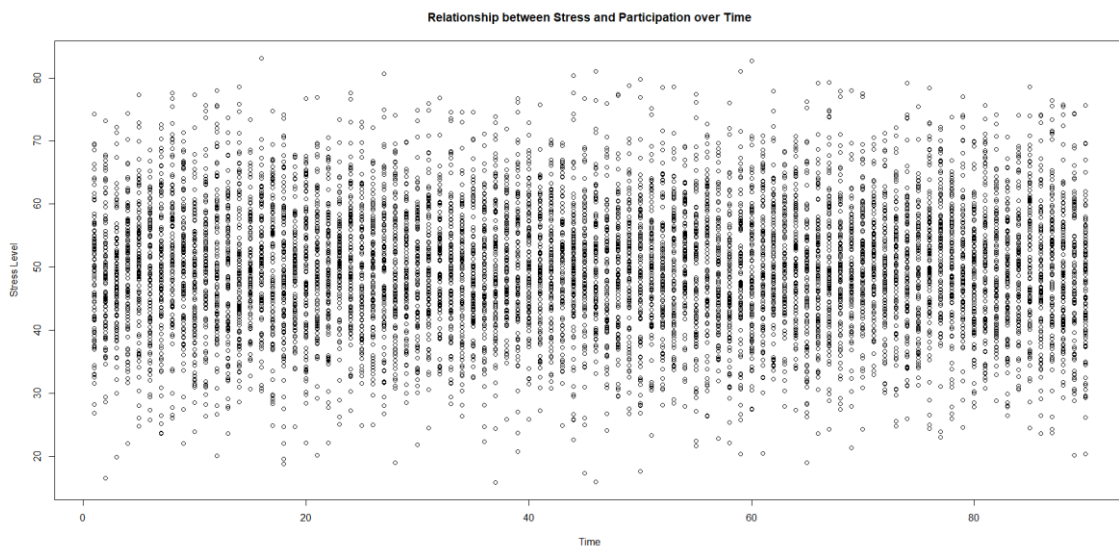


Figure 2: Plot the relationship between stress and participation over time for each participant.

It can be seen that these two variables, at the individual level, at each of the moments, have a distribution, with different variance and mean values, but with comparable behaviors. This can give us indications that we may be facing an ergodic dataset.

After generating the dataset, the code continues with additional calculations on the variables within the dataset. Here's a breakdown of the code:

individual_means <- apply(data[, c(...)], 1, mean): This line calculates the mean for each individual across a selection of variables. The **apply()** function is used to apply the **mean()** function row-wise (**1**) to the specified columns (**c(...)**) in the **data** dataset. These columns include "anxiety", "stress_levels", "genetic_predisposition", and so on. The resulting vector, **individual_means**, stores the calculated means for each individual.

individual_variances <- apply(data[, c(...)], 1, var): Similarly, this line calculates the variance for each individual across the same selection of variables. The **apply()** function with **var()** as the applied function is used to compute the variance row-wise for the

specified columns in the **data** dataset. The resulting vector, **individual_variances**, contains the calculated variances for each individual.

group_mean <- colMeans(data[, c(...)]): This line computes the mean for the entire group across the selected variables. The **colMeans()** function calculates the column-wise mean for the specified columns in the **data** dataset. The resulting vector, **group_mean**, stores the mean values for the entire group.

group_variance <- apply(data[, c(...)], 2, var): Likewise, this line computes the variance for the entire group across the same selection of variables. The **apply()** function with **var()** as the applied function is used to calculate the variance column-wise for the specified columns in the **data** dataset. The resulting vector, **group_variance**, contains the variances for the entire group.

These calculations provide insights into the central tendency (mean) and variability (variance) of the selected variables at both the individual and group levels. By examining individual means and variances, researchers can explore variations in these variables among different individuals. Similarly, the group mean and variance provide a summary of the average values and dispersion across the entire dataset.

These summary statistics can help researchers understand the characteristics and distributions of the variables in the dataset, which can be useful in subsequent analyses or when interpreting the results of survival models or other statistical analyses.

The provided results show the correlation coefficients for the relationships between mean and variance at both the individual and group levels. Here's the interpretation of the results:

- **Individual Mean-Variance Relationship**: The code calculates the correlation between individual means (**individual_means**) and individual variances (**individual_variances**). The correlation coefficient is 0.4910476. This positive correlation suggests a moderate association between the mean and variance of the selected variables at the individual level.

The correlation coefficient ranges from -1 to +1. A positive value indicates that individuals with higher means tend to have higher variances, while individuals with lower means tend to have lower variances. Conversely, a negative correlation would indicate an inverse relationship, where individuals with higher means have lower variances, and vice versa. In this case, the positive correlation suggests that individuals with higher average values for the selected variables tend to exhibit more variability or dispersion in those variables.

- **Group Mean-Variance Relationship**: The code calculates the correlation between the group means (**group_mean**) and group variances (**group_variance**). The correlation coefficient is 0.7636821. This indicates a strong positive correlation between the mean and variance at the group level.

The strong positive correlation implies that as the overall mean value for the selected variables increases, the corresponding variance also tends to increase. In other words, when the group has higher average values, there is greater variability or dispersion within the group.

These correlation coefficients provide insights into the relationship between mean and variance at both individual and group levels. The results suggest that there is a positive

association between the mean and variance of the selected variables, indicating that higher mean values are associated with increased variability at both the individual and group levels. However, it's important to note that correlation does not imply causation, and further analyses or modeling may be required to understand the underlying factors contributing to these relationships.

The presence of ergodicity can be observed in this dataset. Ergodicity implies that the statistical properties of the population can be inferred from a single long-run observation or by analyzing individual observations. In this case, the relationship between anxiety and participation over time is depicted for each participant, showing how their values fluctuate. Additionally, the mean and standard deviation of the anxiety coefficients provide insights into the overall variability within the population.

4. How to manage ergodicity in survival analysis.

Ergodicity analysis can improve survival analysis by allowing researchers to estimate the long-term survival rates of a population more accurately. This can be particularly useful in cases where the survival rates of the population are not constant over time, or where the population is subject to significant changes in the underlying risk factors that affect survival.

One way that ergodicity analysis can improve survival analysis is by allowing researchers to account for changes in the underlying risk factors that affect survival. For example, consider a population of patients with a certain medical condition. If the risk of death for these patients is not constant over time, but rather varies based on factors such as changes in treatment regimens or changes in the patient's overall health status, then ergodicity analysis can be used to account for these changes and more accurately estimate the long-term survival rates of the population.

In another way in which ergodicity analysis can improve survival analysis is by allowing researchers to account for the impact of competing risks on survival. Competing risks refer to the fact that individuals in a population may face multiple potential causes of death, and the probability of death from one cause may depend on the occurrence of death from another cause. Ergodicity analysis can be used to account for competing risks and more accurately estimate the overall survival rates of a population.

Several different formulas and techniques can be used in ergodicity analysis to improve survival analysis. One such technique is the Kaplan-Meier estimate, which is a commonly used method for estimating the survival function of a population. The Kaplan-Meier estimate can be modified to account for changes in the underlying risk factors or competing risks that affect survival, allowing researchers to estimate the long-term survival rates of a population more accurately. Other techniques that can be used in ergodicity analysis include the Cox proportional hazards model and the Aalen-Johansen estimator.

In a survival study, the non-consideration of ergodicity can limit the accuracy of the results because it means that the long-term survival rates of a population cannot be accurately inferred from a single, long-term observation of the population. Instead, it may be necessary to make multiple observations of the population over time to accurately determine its survival rates.

Ergodicity refers to the property of a system or process being statistically representative of its long-term state over time. In the context of dynamical systems theory and

mathematical statistics, ergodicity can be defined in several ways, each with its mathematical formula.

4.1 Ergodicity in Survival in Machine Learning Models

Ergodicity analysis can be used to improve survival analysis using survival machine learning models such as artificial neural networks (ANNs), random forests, and gradient boosting models, to consider some of the most widely used machine learning models. These models can be used to analyse large datasets and identify patterns and trends in the data that may not be visible using other techniques.

For example, ANNs are a type of machine learning model that can be used to analyse survival data and predict survival outcomes. ANNs can be trained on large datasets of survival data and can learn to identify patterns and trends in the data that are associated with improved survival outcomes. By using ANNs in combination with ergodicity analysis, researchers can estimate more accurately the long-term survival rates of a population and identify factors that may be associated with improved survival outcomes.

Similarly, random forests and gradient-boosting models can be used to analyse survival data and predict survival outcomes. These models can be trained on large datasets of survival data and can identify patterns and trends in the data that are associated with improved survival outcomes. By using random forests and gradient boosting models in combination with ergodicity analysis, researchers can more accurately estimate the long-term survival rates of a population and identify factors that may be associated with improved survival outcomes.

But the problem that limits the use of machine learning models for survival analysis is that all of them, when analysing the behaviour of people or organizations, assume ergodicity, often without demonstrating it. This occurs in parametric, semi-parametric, non-parametric and machine learning models ([15] [16]).

Overall, the use of survival machine learning models in combination with ergodicity analysis can help researchers estimate more accurately the long-term survival rates of a population and identify factors that may be associated with improved survival outcomes.

In summary, not considering ergodicity in machine learning models for survival analysis could lead to inaccurate predictions of long-term behaviour and limited generalizability of findings, which could weaken the validity and reliability of the results obtained.

5. Conclusion

The lack of consideration for ergodicity in survival machine learning models poses a relevant weakness. Ergodicity is important in survival analysis because it allows us to draw accurate conclusions at both the group and individual levels. When dealing with non-ergodic sets, the characteristics of the set cannot be used to infer something about a specific individual from that set.

Survival analysis often involves segmenting populations and assuming that a small representative sample can be used to draw conclusions. However, if the dataset is non-ergodic, the results will not coincide, leading to a margin of error in the analysis. This error arises from the non-ergodicity of the set and can impact the accuracy and generalization capabilities of the machine learning models.

By incorporating ergodicity analysis into survival analysis, researchers can improve the estimation of long-term survival rates, especially when dealing with varying risk factors

and competing risks. Ergodicity analysis allows for the accurate estimation of survival rates over time, accounting for changes in the underlying factors affecting survival.

We have shown that if, in the process of designing the experiment, the data collection is correctly planned in such a way that it allows us to have a longitudinal and continuous vision during the time of each agent (person, company, machine, etc.), the ergodicity analysis is relatively easy to test.

Survival machine learning models such as artificial neural networks, random forests, and gradient boosting models can benefit from ergodicity analysis to enhance their predictive capabilities. These models can identify patterns and trends in large datasets that are associated with improved survival outcomes. By combining ergodicity analysis with these models, researchers can estimate survival rates more accurately and identify factors that contribute to better survival outcomes.

However, a limitation of machine learning models for survival analysis is the assumption of ergodicity without proper demonstration. This assumption is present in parametric, semi-parametric, non-parametric, and machine learning models, and it can weaken the validity and reliability of the results obtained.

References

- [1] P. C. Molenaar, «Consequences of the ergodic theorems for classical test theory, factor analysis, and the analysis of developmental processes,» *Handbook of cognitive aging*, pp. 90--104, 2008.
- [2] Pearl, Judea and others, *Models, reasoning and inference*, Cambridge, UK: Cambridge University Press, 2000.
- [3] W. S. Robinson, «Ecological correlations and the behavior of individuals,» *International journal of epidemiology*, pp. 337--341, 2009.
- [4] Hamaker, E. L., Dolan, C. V., & Molenaar, P. C., «Statistical modeling of the individual: Rationale and application of multivariate stationary time series analysis,» *Multivariate behavioral research*, vol. 40, n° 2, pp. 207--233, 2005.
- [5] Hamaker, E. L., Nesselroade, J. R., & Molenaar, P. C., «The integrated trait--state model,» *Journal of research in personality*, vol. 41, n° 2, pp. 295--315, 2007.
- [6] Molenaar, Peter CM and Campbell, Cynthia G, «The new person-specific paradigm in psychology,» *Current directions in psychological science*, pp. 112--117, 2009.
- [7] Beltz, Adriene M and Wright, Aidan GC and Sprague, Briana N and Molenaar, Peter CM, «Bridging the nomothetic and idiographic approaches to the analysis of clinical data,» *Assessment*, pp. 447--458, 2016.
- [8] A. J. Fisher, «Toward a dynamic model of psychological assessment: Implications for personalized care.,» *Journal of consulting and clinical psychology*, p. 825, 2015.
- [9] Fisher, Aaron J and Medaglia, John D and Jeronimus, Bertus F, «Lack of group-to-individual generalizability is a threat to human subjects research,» *Proceedings of the National Academy of Sciences*, pp. E6106--E6115, 2018.
- [10] O. Peters, «The ergodicity problem in economics,» *Nature Physics*, pp. 1216--1221, 2019.
- [11] Doctor, Jason N and Wakker, Peter P and Wang, Tong V, «Economists' views on the ergodicity problem,» *Nature Physics*, pp. 1168--1168, 2020.
- [12] Borkenau, Peter and Ostendorf, Fritz, «The Big Five as states: How useful is the five-factor model to describe intraindividual variations over time?,» *Journal of Research in Personality*, pp. 202--221, 1998.
- [13] Na, Jinkyung and Grossmann, Igor and Varnum, Michael EW and Kitayama, Shinobu and Gonzalez, Richard and Nisbett, Richard E, «Cultural differences are not always reducible to individual differences,» *Proceedings of the National Academy of Sciences*, pp. 6192--6197, 2010.
- [14] Molenaar, P., Sinclair, K. O., Rovine, M. J., Ram, N., & Corneal, S. E., «Analyzing developmental processes on an individual level using nonstationary time series modeling,» *Developmental psychology*, vol. 45, n° 1, p. 260, 2009.
- [15] Adolf, Janne K and Fried, Eiko I, «Ergodicity is sufficient but not necessary for group-to-individual generalizability,» *Proceedings of the National Academy of Sciences*, pp. 6540--6541, 2019.

- [16] Mangalam, M., & Kelty-Stephen, D. G., «Ergodic descriptors of non-ergodic stochastic processes,» *Journal of the Royal Society Interface*, vol. 19, n° 189, p. 20220095, 2022.
- [17] E. H. Simpson, «The interpretation of interaction in contingency tables,» *Journal of the Royal Statistical Society: Series B (Methodological)*, pp. 238--241, 1951.
- [18] Berestycki, N and Etheridge, AM and Hutzenthaler, M, «Survival, extinction and ergodicity in a spatially continuous population model,» *Markov process. related fields*, pp. 265--288, 2009.
- [19] Heino, Matti TJ and Knittle, Keegan and Noone, Chris and Hasselman, Fred and Hankonen, Nelli, «Studying behaviour change mechanisms under complexity,» *Behavioral Sciences*, p. 77, 2021.
- [20] Li, Qiuya and Zhao, Dianli, «Survival and ergodicity of a stochastic microorganism flocculation model with nonlinear response functionals,» *Nonlinear Dynamics*, pp. 2663--2680, 2023.
- [21] Lowie, Wander M and Verspoor, Marjolijn H, «Individual differences and the ergodicity problem,» *Language Learning*, pp. 184--206, 2019.
- [22] Norris, Lesley A and Kendall, Philip C, «Moderators of outcome for youth anxiety treatments: Current findings and future directions,» *Journal of Clinical Child & Adolescent Psychology*, pp. 450--463, 2021.