Analysis of R-R Intervals Using Point Process and Time Series Methods

By: Reetom Gangopadhyay

MA568, Professor Eden
ID: U76940851

Abstract:

This paper seeks to investigate cardiovascular and autonomic regulation through a tilt-table protocol. The data contains four healthy subjects who were a part of the tilt-table experiment under two separate conditions. The following analysis will attempt to answer questions about structural changes under specific conditions, how the heart adapts to each condition, and whether there are predictive capabilities for this data using several statistical methods. To answer these questions, I employ visualizations, several point process models, and time series models. The results will show the importance of history dependence/lags, highlight changes in firing patterns during the change of positions, and compare the predictive capabilities of each type of model.

Introduction:

Cardiovascular health is paramount when speaking to the overall well-being of an individual; therefore, cardiovascular health is one of the most pronounced aspects among the traits of an overall healthy individual. Autonomic regulation and responsiveness to changes in body position are intrinsically linked to an individual's cardiovascular health. The tilt-table protocol serves as a valuable framework to experiment and study the interactions resulting from the procedure. We observe R-R intervals, the time elapsed between two R waves from an ECG/EKG and may use this data to evaluate the relationship between cardiovascular and autonomic systems.

Using the tilt-table protocol, we may uncover insights into the adaptive behavior of the cardiovascular system, mainly focusing on the R-R intervals. By understanding autonomic regulation, doctors may be able to identify medical conditions more promptly and more accurately. The implications of a more prompt and accurate identification of a disorder in the heart can lead to vastly different outcomes for a patient.

By exploring visual tools to gauge the approximate structure of the data and then employing several point process models such as the Poisson or inhomogeneous Poisson models, inverse Gaussian models, and an ARIMA model, we can approach the problem of attempting to estimate the dynamic system modeled in the experiment.

Despite the established metrics, it is imperative to continue research as the definitions and general modeling metrics have evolved over time. This study approaches the problem using two separate lenses, one using a more traditional spiking/point-process modeling criteria and another using more applied time series modeling and decompositions. Using such a diverse statistical toolkit allows a more nuanced view of the problem of heart rate and variability in the context of a tilting experiment.

This paper seeks to connect the technical intricacies of a statistical model and the broader, more general implications for human health. The implications of research into the tilt-table experiment can have profound impacts on the world of medicine, including, but not limited to, clinical diagnostics, more personalized medical care, and, in a more general view, our understanding of

physiological adaptation. Applying thoughtful statistical insights to a multifaceted problem, such as the test mentioned earlier, gives us a far greater intuition on how to make future health decisions.

Dataset:

The dataset comprises heartbeat data from four healthy subjects under two conditions (a and b). In condition a, the subject starts in the horizontal position and, at 180 seconds, is tilted to the vertical position. In condition b, the subject starts in the horizontal position and, at 180 seconds, will stand up by themselves. Each dataset (1a/b, 2a/b, 3a/b, 4a/b) contains the R-R intervals and the time of the event. The data can be represented as a time series.

Exploratory Data Analysis:

The heartbeat data from the EKG provides a record of the R-R intervals. Preliminary data analysis consisted of a variety of visualization tools, including time series plots, histograms of the spike times, and a density plot showing the distribution of the recorded R-R intervals.

The time series plots show a general trend of the R-R intervals over the course of the experiment. These plots are imperative to view the overall fluctuations in heart rate and variability as subjects transitioned from horizontal to vertical states or vice versa. Plots were made for every subject in every condition.

Histograms of the spiking activity give a perspective on the distribution of the R-R intervals. The binning shows a general distribution of the spike times and will also show the dominant times and perhaps other important features. Histograms were created for all conditions.

Finally, a density plot for each condition was created to give a continuous representation of the distributions. The smoothing will assist in identifying a general distribution.

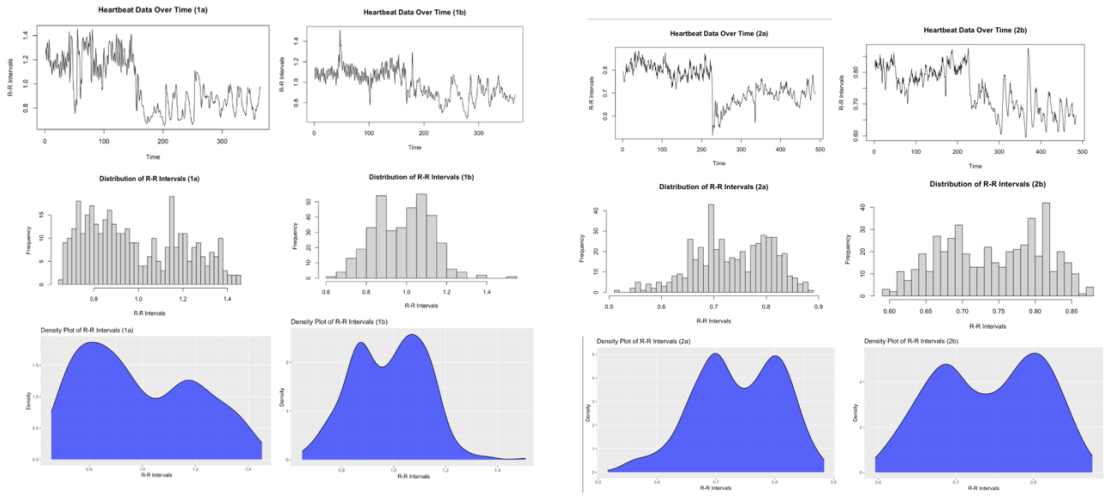


Fig: Visuals for subjects 1 and 2 (for both, a on left, b on right)

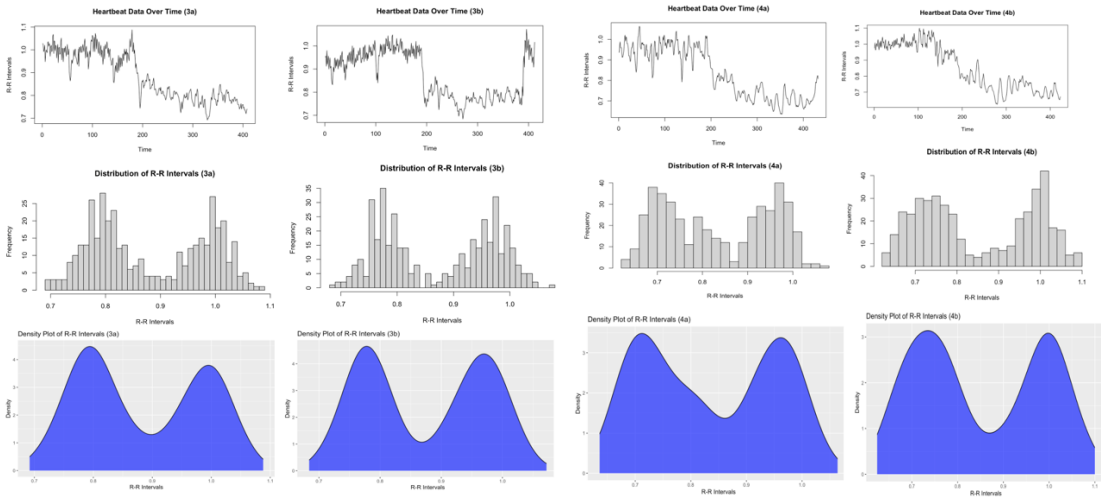


Fig: Visuals for subject 3 and 4 (for both, a on left, b on right)

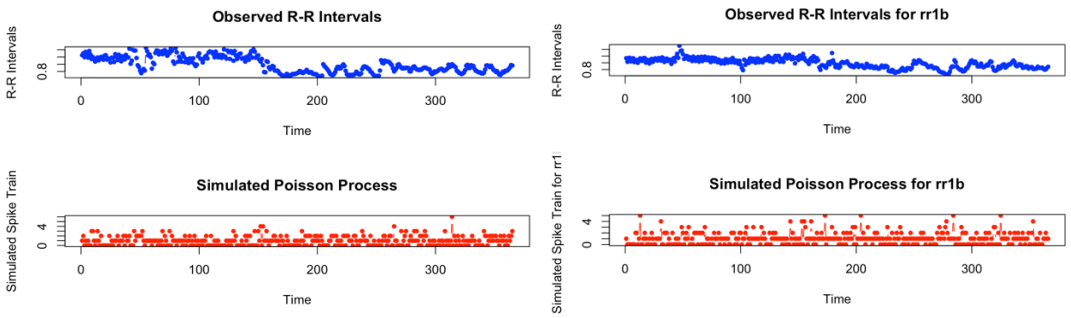
The visuals show a noticeable structural break in the R-R intervals at the 180-second mark (as mentioned in the problem) for all subjects in both conditions. Each model also shows bimodality. However, some subjects have a higher valley between the peaks than others. When approaching the problem through the lens of time series, for stationarity consideration, we should expect to difference the data to fit a model. A form of Generalized Linear Model (GLM) may fit the behavior of the data well.

Methods/Results:

For brevity, the analysis presented will be for 1a/b, but each method can be successfully applied to any of the other distributions.

Simple Poisson Process:

As a preliminary modeling method, we may attempt to fit the simple Poisson process to the data. The Poisson process is simple to implement and very easily interpretable, generally providing parsimonious models. The Poisson process is a stochastic model that characterizes the occurrence of each event in a continuous time. In the context of the heartbeat dataset, each R-R interval is an event, and the process will assume independence of intervals with a constant rate. Fundamentally, the Poisson process assumes that the probability of observing events in the given interval is discretized by a Poisson distribution. While an attractive starting point for modeling, the basic Poisson process tends to fall short when handling more intricate processes. More complicated analyses, such as the inhomogeneous Poisson model, may be beneficial to enhance model efficacy.

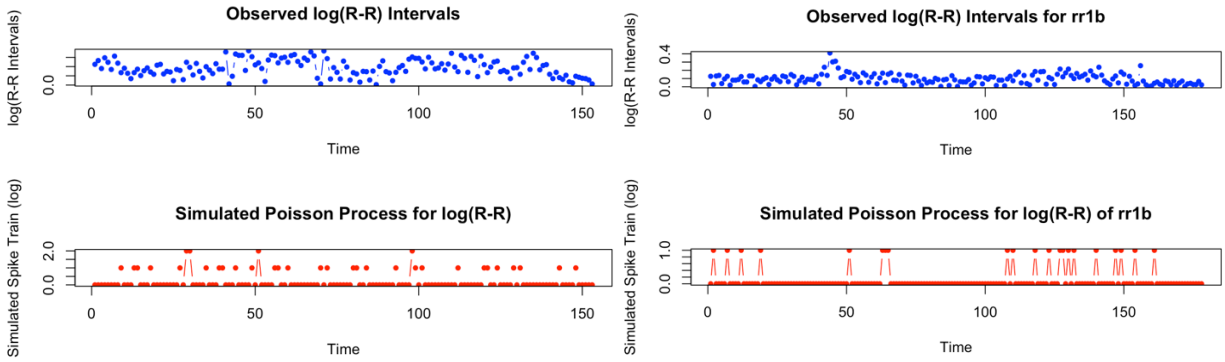


Left Fig: Simple Poisson Process 1a
Right Fig: Simple Poisson Process for 1b

To assess model efficacy, we perform the Kolmogorov–Smirnov test. The KS test is defined as follows:
H_0: The observed data and the simulated process are from the same continuous distribution
H_A: The observed data and simulated process are NOT from the same continuous distribution

In this case, the p-value is < 2.2e-16 for both a and b. This tells us that the observed R-R intervals are not from the same distribution as the simulated process. The patterns also look nothing alike in the plots above.

To remedy this issue, we may perform a log transform to the R-R intervals.

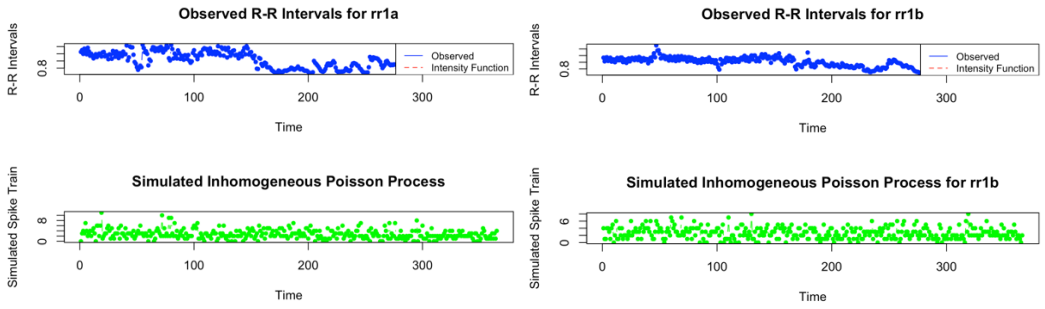


Left Fig: Simple Poisson Process for Log(R-R), 1a
Right Fig: Simple Poisson Process for Log(R-R), 1b

Generally, the fit appears to look better. The spikes are captured more accurately by the log transform. However, the fit is still not accurate enough as shown by the KS test once again giving the p-values of 4.746e-09 for 1a and 0.0001231 for 1b. This means that the observed data is still not generated by the Poisson process for log(R-R).

Inhomogeneous Poisson Process:

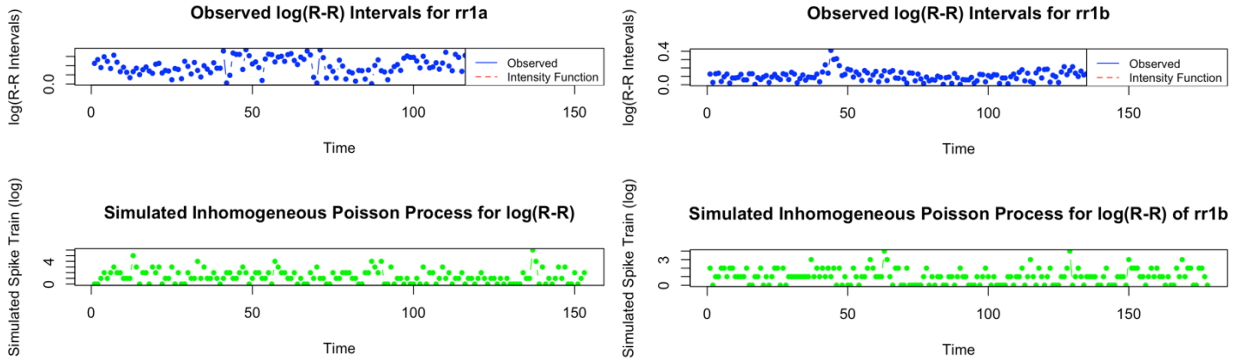
Next, we may attempt to fit an inhomogeneous Poisson model. Building off the simplicity of the Poisson process, we may extend our point-process modeling toolkit with an inhomogeneous Poisson process. This process rejects the assumption of a constant rate of the simple Poisson process. Thus, the Poisson process inhomogeneous allows for a time-varying rate, making it more effective for dynamic changes. In our case, because the heartbeat intervals change, it may perhaps be a better fit.



Left Fig: Inhomogeneous Poisson Process for (R-R), 1a
Right Fig: Inhomogeneous Poisson Process for (R-R), 1b

The p-values for the KS test in this case are both < 2.2e-16. This means we, once again, cannot reject the null hypothesis of the KS test. Visually, the simulated processes do not capture the spiking properties of the observed data.

Like the simple Poisson process, we may attempt a log transformation for this process as well.



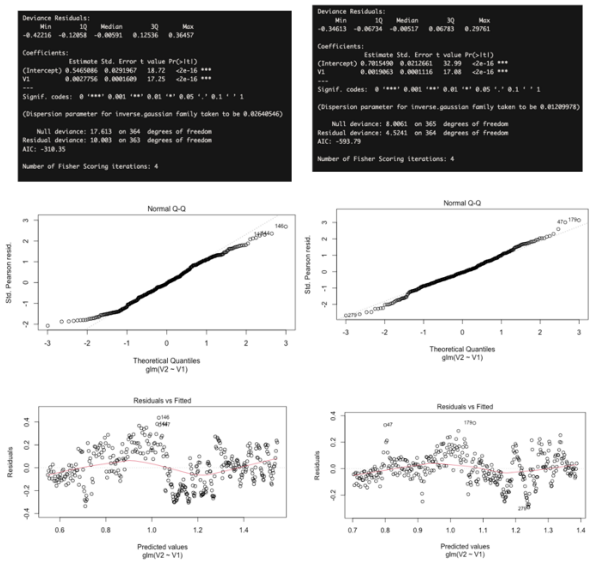
Left Fig: Inhomogeneous Poisson Process for log(R-R), 1a
Right Fig: Inhomogeneous Poisson Process for log(R-R), 1b

The P-values for the KS test are < 2.2e-16 for both processes again. This tells us that there was no appreciable benefit to using the log transform for either case.

Inverse Gaussian GLM:

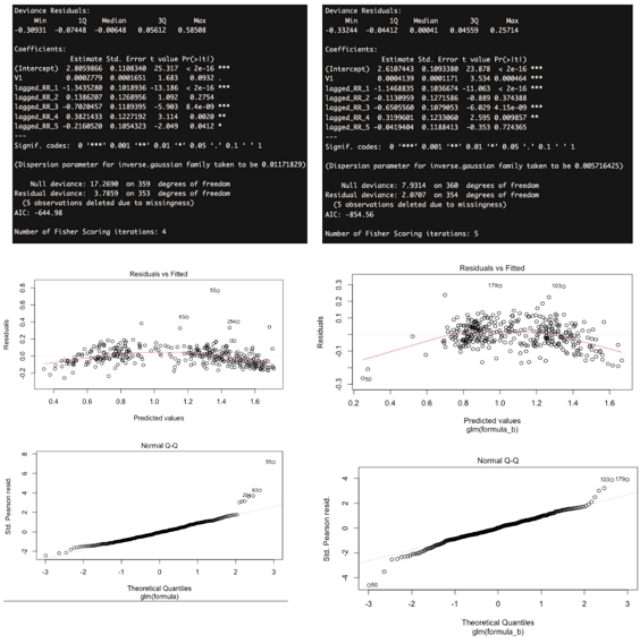
In the paper “A point-process model of human heartbeat intervals: new definitions of heart rate and heart rate variability” by Barbieri, Matten, et al. (2004), the authors suggest the usage of a point process method utilizing the inverse Gaussian distribution, and that the spiking activity from this heartbeat dataset should fit this process well. To expand on their work, using spiking history may be helpful.

To begin fitting this model we should use a Generalized Linear Model (GLM). Using a GLM gives a more sophisticated understanding of heartbeat intervals, as they extend the traditional linear regression framework to accommodate variables with a non-Gaussian distribution. Since our data distribution is not Gaussian as shown by the visualizations, we expect this approach to be beneficial in our context. The inverse Gaussian distribution within the GLM is ideal as this distribution is a strong predictor for positively skewed data and handles variability in the response variable well. This is ideal for the heartbeat data as there is a noticeable structural break. To fit this model, we should use the R command glm() and set the parameter “family” to inverse.gaussian with the link function 1/mu^2.



Top: Readout of GLM (Left A, Right B)
Middle: QQ Plot (Left A, Right B)
Bottom: Residual plot (Left A, Right B)

The readouts show a low residual deviance, and the parameter estimates are significant, leading us to believe that this model is a reasonably good fit. The QQ plot shows ideal behavior for a non-Gaussian distribution for both processes. However, the residual plot still shows some pattern in both. To address the pattern in the residuals, we may use history dependence and create lags as a parameter for the GLM. History dependence is the technique implementing the creation of lags as extra model parameters including the previous value for the spike interval.



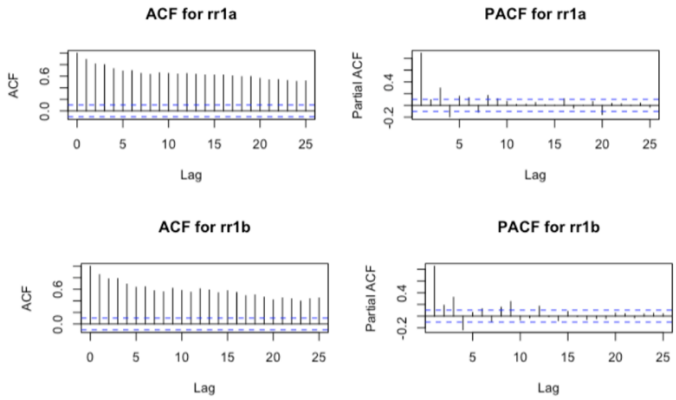
Left: 1a, Right: 1b

The residual plots still show patterns, but this largely appears to be a result of certain outliers creating a pattern when the overall pattern may have otherwise been random. The lags appear to have helped the model fit. As shown by the p values, the lags are generally significant except for lag 2 for both 1a and 1b. The process is likely history-dependent and benefits from having prior knowledge. This means that previous R-R intervals will influence the next intervals. Overall, we can say that this model is a good fit for the data.

Time Series:

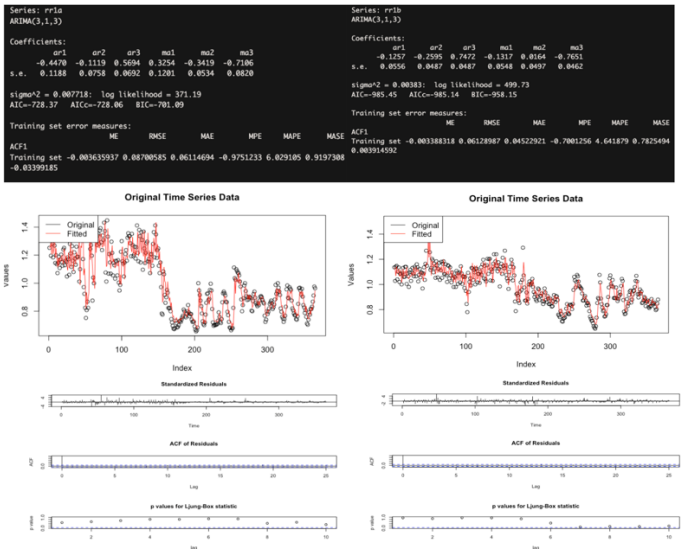
In a separate approach to modeling, we may use time series modeling methods such as the ARIMA model and Spectral decomposition to attempt to capture the spiking pattern.

Due to the structural break in the data that induces nonstationarity, we must difference the data when fitting a model. This is shown more clearly by the ACF/PACF plots.



Left: 1a, Right: 1b

The plots show evidence of a unit root due to the long decay times of the ACF values. Using the auto.arima() function in R, we can fit the ideal ARIMA model using the minimum AIC values.



Left: 1a, Right: 1b

For both sets of data, the model fit is ARIMA(3,1,3), which is an ARIMA model 1st order differenced with an ARMA(3,3) hybrid process. Visually, the fitted model strongly indicates how the model will move over time. The statistics included also illustrate the same conclusion. The Standardized residuals do not have any significant outliers for either 1a or 1b, the differenced ACF also appears to have no significant lags, and the Ljung-Box statistic has high p-values for all lags. We can see that the general change in heartrate when the subject stands by themselves is lower than when they are tilted in experiment A. The Ljung-Box statistic is designed to show whether there is significant autocorrelation at any or all lags. The hypotheses are as follows:

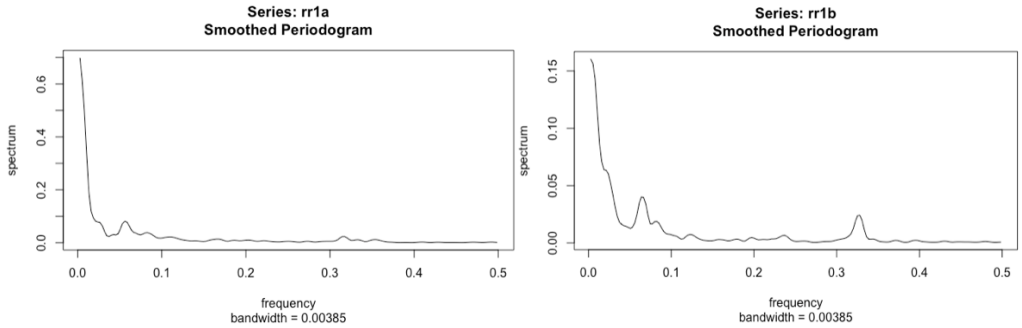
H_0: There is no autocorrelation in the data at different lags.
H_A: There is significant autocorrelation in the data at one or more lags.

Due to the high p-values, we cannot reject the null hypothesis at any lag, which means that the ARIMA process is likely to be a strong fit to the data. The ACF of the residuals decaying quickly means that the differencing carried out in the ARIMA model was effective.

Generally, we conclude that the ARIMA model is a good fit for this data.

Spectral Analysis:

Spectral analysis is a powerful technique used to explore the dominant frequency components embedded within a time series. In the context of the heartbeat intervals, we may explore periodic patterns and frequency changes in cardiovascular response. To start, we should look at the spectral density plot. If there are dominant frequencies in the data, we can use them in tandem with the aforementioned models to boost model efficacy.



Left: 1a, Right: 1b

Based on the spectral density plot, there does not appear to be any dominant frequencies, but a slight spike occurs at around 0.35 on both plots. The spike is more pronounced on the 1b plot. The periodicity is around 2.8 (1/0.35) if this is a significant frequency, meaning the period is completed after roughly 3 cycles. However, it appears that the peaks at these frequencies are relatively minor. Therefore, inferential methods may be used to test a hypothesis to judge if these peaks are significant.

Discussion/Conclusion:

As previously mentioned, a comprehensive modeling approach gives a more holistic view of the data.

Point process models, including simple and inhomogeneous Poisson processes, provide insights into the dynamics of heartbeat intervals. The simplicity of a Poisson process captured some basic patterns, while the inhomogeneous Poisson process was tested to see if time-varying rates existed within the data. Moving to a GLM with the inverse Gaussian Distribution gave a more sophisticated approach to modeling the intervals, considering non-gaussian data, and certainly captured the variability better than the previous modeling methods.

Using time series models allowed a different ideology to give a more comprehensive look into the patterns generated by the spiking activity. The ARIMA model successfully fits the spiking pattern well, showing the pattern of the spikes with very high accuracy. The ARIMA model is easily interpretable, and the results are simple. Employing spectral analysis and using spectral density unveiled potential frequency components and rhythmic patterns in the intervals. Accounting for dominant frequencies in the data allows for more complex modeling methods to be explored in the future.

The Poisson processes are not advanced enough to capture the more complicated pattern of the heartbeat data. They assume independence between events within the small-time intervals, which may oversimplify some inherent data characteristics. One way to fix this would be to remove observations that are too close to each other to attempt to enhance the model.

The inverse Gaussian GLM model can be challenging to interpret when factoring in the spiking history. It is crucial to assess the underlying assumptions carefully to ensure a proper fit. The inverse Gaussian model can also be used for right-skewed data; the density plots are shown before demonstrating bimodality. In the paper by Barbieri, Matten et al. (2004), they fail to create a density plot and check for bimodality; the model still appears to fit the spiking pattern well in the paper. Therefore, it is possible to ignore this characteristic, perform the analysis with this model type, and still achieve reasonable results. Future research may include using a Gaussian Mixture Model to tackle the bimodal distribution directly. Alternatively, considering the structural change, we may be able to create two different GLMs in an ensemble with each other. The first would be before the 180-second mark, the second of which would be after the 180-second mark. Using these two models together could potentially give a more robust result.

Expansion of the ARIMA model to more complicated methods, such as ARCH or GARCH errors, may help capture sub-processes in the data and provide a more precise fit. In addition, one can also consider a time series model that allows for structural breaks, which may be appropriate for this situation. It is vital to assess the interpretation of the time series model against the point process model as they may sometimes have different or conflicting results. Spectral analysis would help improve model accuracy as well.

In conclusion, the multi-faceted approach outlined in this paper to analyze heartbeat data has provided a nuanced understanding of the R-R intervals and general adaptability of the cardiovascular system given stimuli. While each method contributed valuable insights, further refinements and improvements could be made. The ensemble of statistical techniques and visualizations offers a complete perspective, giving a solid foundation to perform additional research in the field of autonomic response. Analyzing these patterns gives medical researchers a base to identify new ways to diagnose disease and disorders in patients.

Works Cited:

Barbieri, Riccardo et al. “A point-process model of human heartbeat intervals: new definitions of heart rate and heart rate variability.” *American journal of physiology. Heart and circulatory physiology* vol. 288,1 (2005): H424-35. doi:10.1152/ajpheart.00482.2003

Roth, Thomas. “RR Interval.” RR Interval - an Overview | ScienceDirect Topics, 2011, www.sciencedirect.com/topics/nursing-and-health-professions/rr-interval#:~:text=RR%20interval%2C%20the%20time%20elapsed,as%20well%20as%20autonomic%20influences.