Objective:

The goal of this lab is to identify groups of patients with similar event dynamics and to model the temporal structure of their medical events using Hawkes processes. Comparing the Hawkes parameters across clusters provides insight into different underlying patient behaviors and potential disease progression patterns.
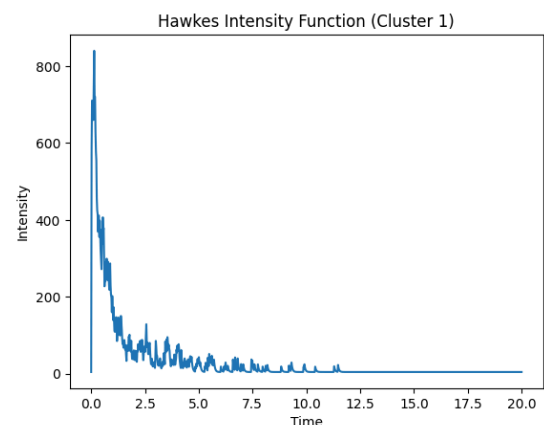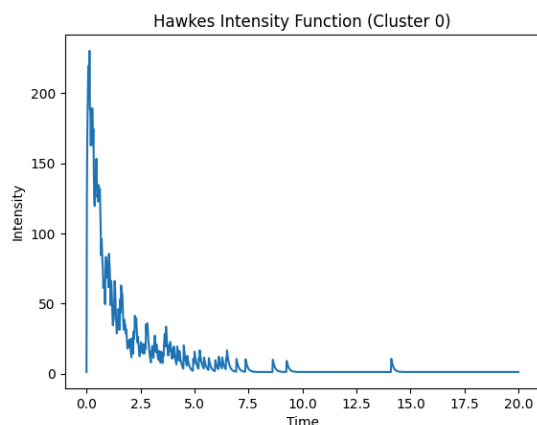
Clustering:

The patient features were continuously scaled using MinMaxScaler. Several values of k were tested, and two clusters were selected as being optimal from their silhouette score. K-Means clustering was then applied and visualized using PCA. Patients were separated into distinct groups with differing clinical feature profiles (e.g., one cluster with higher F0–F2 scores indicating more severe baseline conditions). The cluster sizes weren't very balanced, with cluster 1 containing 855 members (681 events) while cluster 0 contained 257 members (245 events).
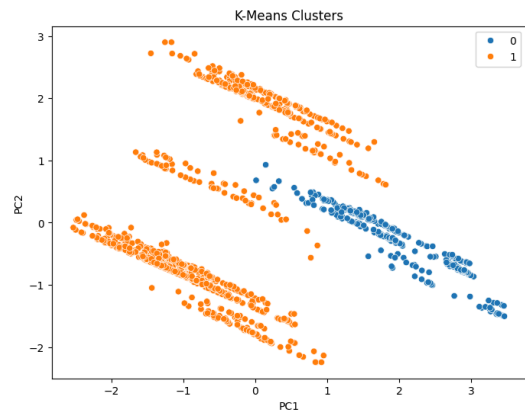
Hawkes Process:

For each cluster, event sequences were modeled using an exponential Hawkes process using the HawkesLib library. The key parameters estimated were μ (baseline intensity), α (excitation amplitude), and β (decay rate). Cluster-level Hawkes parameters were compared visually to interpret differences in event dynamics between patient groups. Cluster 1 exhibits a high baseline intensity (μ = 4.55) and a fast decay (β = 21.87) while cluster 0 shows a lower baseline intensity (μ = 1.22) but stronger persistence (lower β = 10.09). Both clusters have similar excitation strength (α ≈ 0.93).

<u>Key Takeaways</u>

Clustering:
- Cluster 1 (n = 855): characterized by moderate baseline feature values, representing the majority of patients with relatively stable profiles.
- Cluster 0 (n = 257): showed higher feature variability, suggesting patients with more severe or complex baseline conditions.

Hawkes Process:
- Cluster 1 exhibited a higher baseline intensity and a fast decay, indicating a relatively steady occurrence of events with limited self-excitation.
  - Patients in this group tend to experience events more regularly and independently, consistent with a stable or chronic progression pattern.
- Cluster 0 shows a lower baseline intensity but a stronger persistence, indicating few spontaneous events that likely trigger a prolonged sequence of additional events
  - Patients in this group tend to be stable but when they do experience a cascade of events, they take longer to settle down, consistent with relapsing or flare-up conditions.
- The difference in Hawkes parameters captures contrasting underlying dynamics.

Conclusion:
Clustering patients by features followed by Hawkes process modeling revealed two distinct subpopulations with different event dynamics. The first cluster represents patients with consistent, independent event patterns, while the second displays more reactive, self-exciting behavior. These insights illustrate how temporal point process modeling can enhance patient stratification and support data-driven decisions in longitudinal clinical studies.

Google Colab:
https://colab.research.google.com/drive/1jdw5kvN4V9hx2YagNSiMI1R58_HYIh5n?usp=sharing

Github:
https://github.com/mryeazel-729/MLHealth/blob/main/Lab2.ipynb