

Evidence Chain

University of California, Irvine

May 9, 2024



Sections

1 Dataset Characteristics

2 Hypothesis Testing



Dataset Characteristics



Research Question

- Research questions: How do people decide to stop
 - Task: In the experiment, +1 means presenting one stimulus (stimulus "O"), and -1 means presenting the other stimulus (stimulus "X").

On each trial, for $n \in N+$, let $\{X_n\}$ be a stochastic sequence where $X_0 = 0$ and $X_n \sim Bern(p)$ if $X_n = \pm 1$, with $\mathbb{P}(X_n = 1) = p$ and $\mathbb{P}(X_n = -1) = 1 - p = q$.

We fixed $p = 0.62$ to make the Bernoulli trials biased. To counterbalance the stimulus, half of the trial sequences are drawn have $\mathbb{P}(X_n = 1) = 0.62$, where stimulus "O" is more likely to be presented, and half of the trial sequences are drawn from $\mathbb{P}(X_n = -1) = 0.62$ where stimulus "X" is more likely to be presented. After training sessions, human subjects need to decide whether the chain is "X" dominant or "O" dominant. The sequence is terminated once a decision is made. Subjects need to make a decision on whether "X" is dominant or "O" is dominant as the sequence is unfolding.

- It's a decision associated with
 - how much evidence they have seen
 - how many samples are presented
 - sequence in which evidence is presented



Formally Defining the Problem

Given the stochastic sequence of iid rv's, $\{X_n\} \in \mathbb{N}^+$, A random variable T taking values in \mathbb{N}^+ is considered to be a stopping time with respect to $\{X_n\}$ if for each $n \in \mathbb{N}^+$ there exists a decision function $G_n : \{-1, 1\}, n \rightarrow \{0, 1\}$, such that:

$$\mathbf{1}\{T = n\} = G_n(X_0, X_1, \dots, X_n), \text{ for all } n \in \mathbb{N}^+ \quad (1)$$

G_n can be thought of as a black box which takes the values of the process $\{X_n\} \in \mathbb{N}^+$ observed up to the present point and outputs either 0 or 1. The value 0 means keep going and 1 means stop.
We want to find out what the G_n decision functions are.



Evaluation Framework

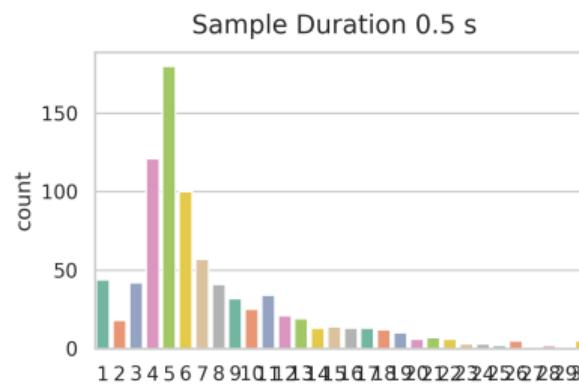
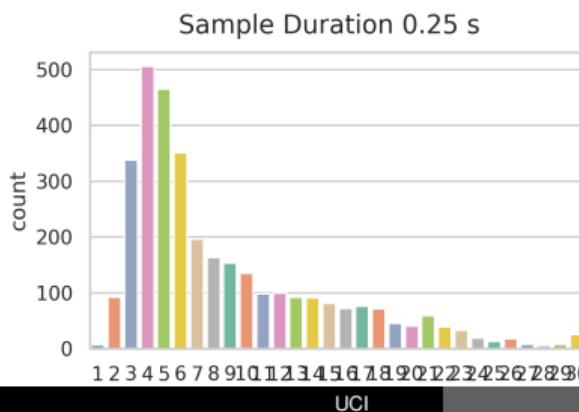
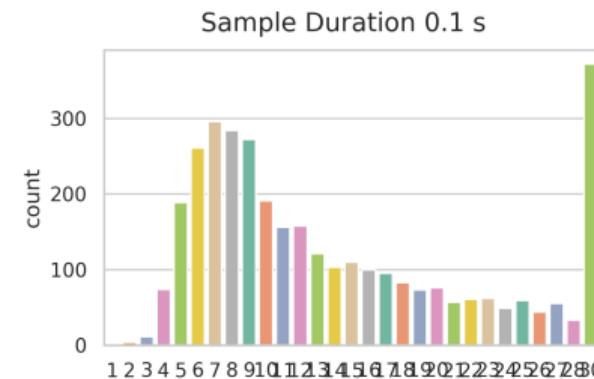
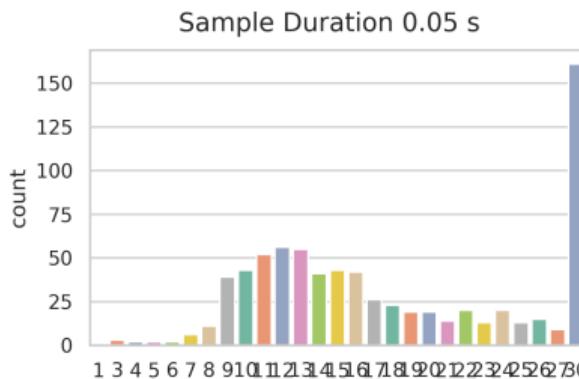
x_1 lock



Accuracy by subjects



Number of Trials at Termination



Percentage of Trials Terminated



Hypothesis Testing



Hypothesis 1: Random Walk - fixed bound at all times vs. bound as a function of samples

Under this hypothesis, we assume that integrate evidence, such that $S_n = \sum_{i=1}^n X_i$, and S_n is therefore a random walk. In other words, the level of evidence would be the difference between positive and negative samples.

- H_0 : The stopping criteria is deterministic, a fixed bound b_0 is used regardless number of samples.

$$G_n(X_0, X_1, \dots, X_n) = \begin{cases} 1, & S_n = b_0 \\ 0, & S_n \neq b_0 \end{cases} \quad (2)$$

- H_1 : The stopping criteria changes as the number of samples increase, where $f(S_n)$ is a function of S_n and n

$$G_n(X_0, X_1, \dots, X_n) = \begin{cases} 1, & S_n = f(S_n, n) \\ 0, & S_n \neq f(S_n, n) \end{cases} \quad (3)$$

This Random Walk hypothesis can also be seen as a *LLR* test.

jenny TODO: 1: insert my proof next page

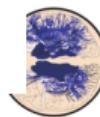
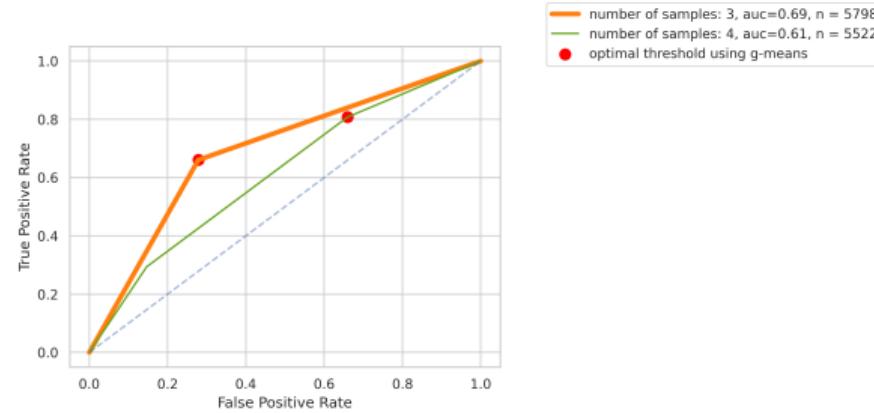
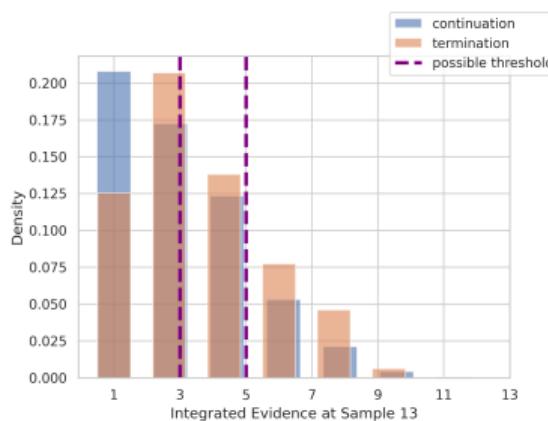


Hypothesis 1: Random Walk - fixed bound at all times vs. bound as a function of samples

To test the hypothesis, with different sample length, we use $S_n = \sum_{i=1}^n X_i$ to classify whether subjects stop or proceed. And evaluate the performance using AUC score (from TPR and FPR at different threshold). We then use geometric means to calculate the optimal threshold.

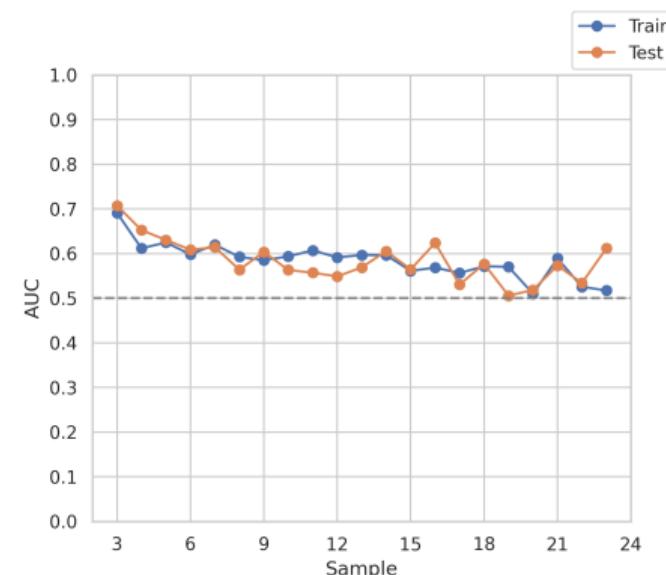
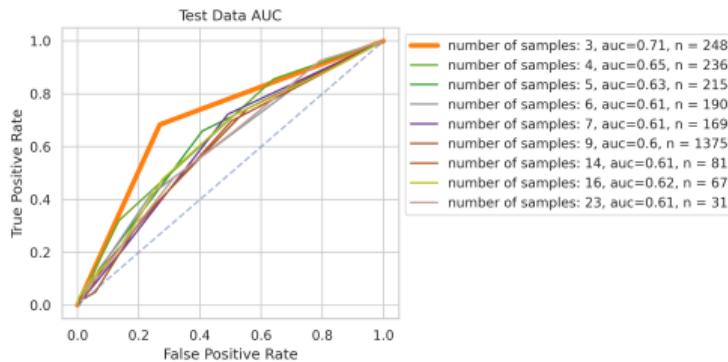
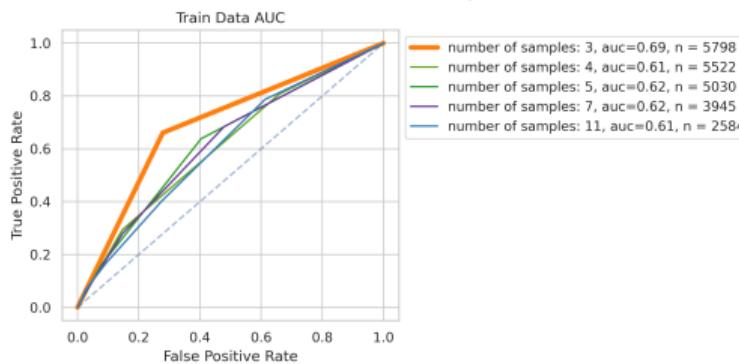
- input = S_n at different sample
 - label = True if the chain is terminated: false if the chain continues

$$GSS = \sqrt{\text{sensitivity} * \text{specificity}} = \sqrt{(TPR * (1 - FPR))} \quad (4)$$



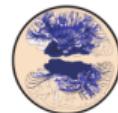
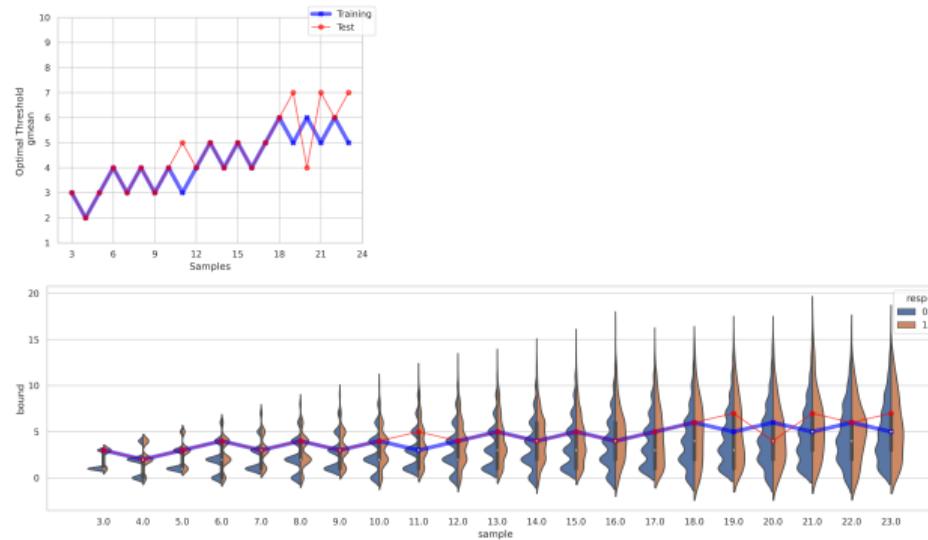
Hypothesis 1: Random Walk - fixed bound at all times vs. bound as a function of samples

All stimulus duration combined (50ms, 100ms, 250ms, 500ms)



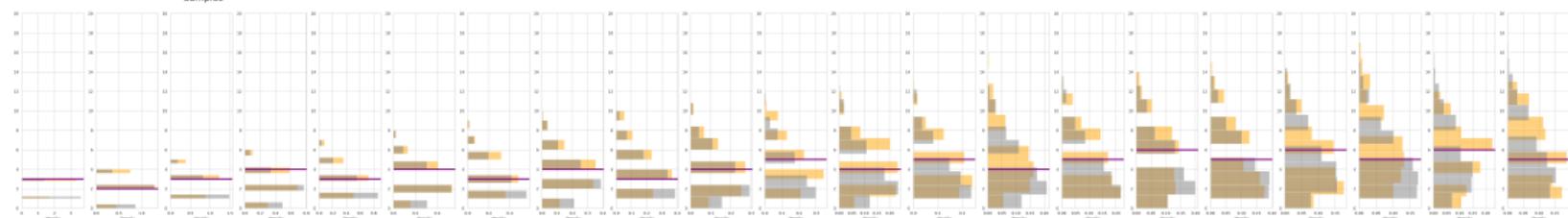
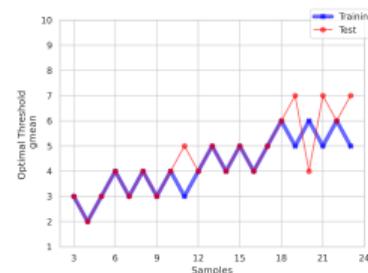
Hypothesis 1: Random Walk - fixed bound at all times vs. bound as a function of samples

All stimulus duration combined (50ms, 100ms, 250ms, 500ms)



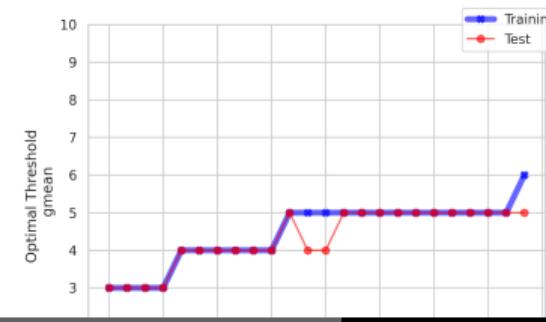
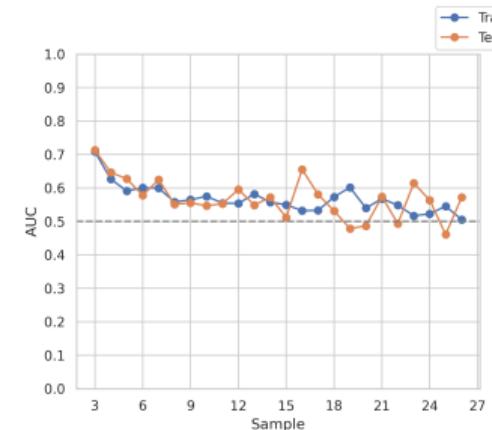
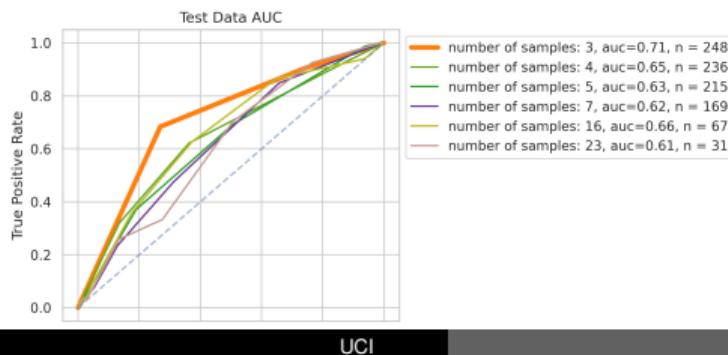
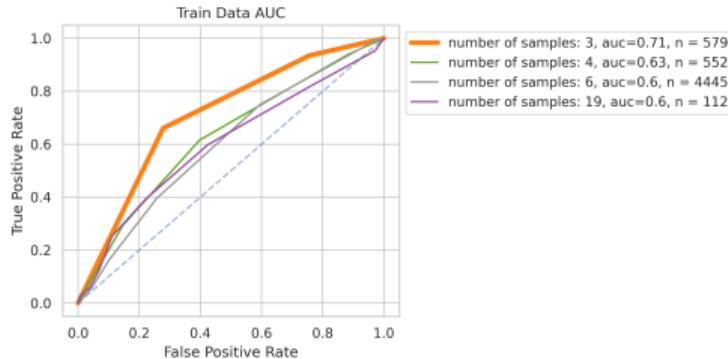
Hypothesis 1: Random Walk - fixed bound at all times vs. bound as a function of samples

All stimulus duration combined (50ms, 100ms, 250ms, 500ms)



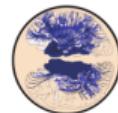
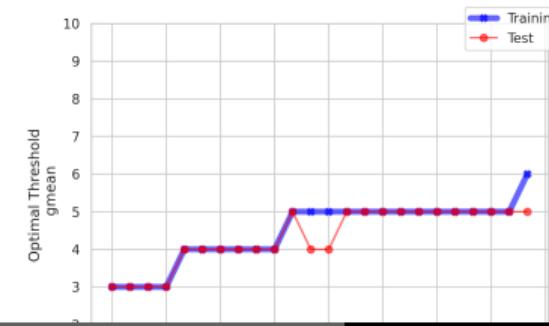
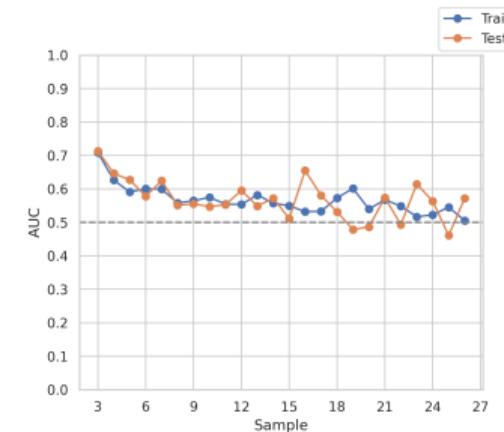
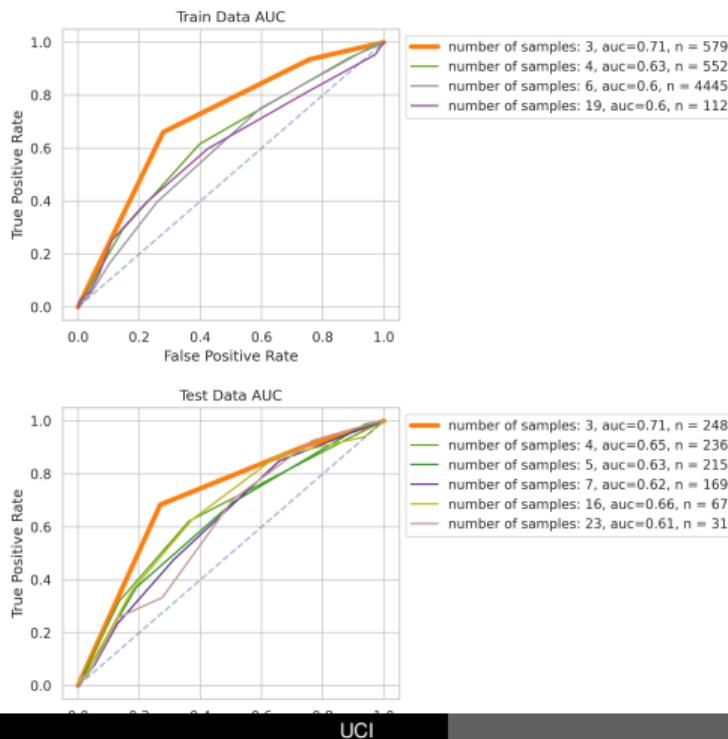
Hypothesis 2: Runs Model- fixed runs at all times vs. runs as a function of samples

All stimulus duration combined (50ms, 100ms, 250ms, 500ms)



Hypothesis 2: Runs Model- fixed runs at all times vs. runs as a function of samples

250ms



Hypothesis 3: Weighted Sum of Random Walk - Logistic

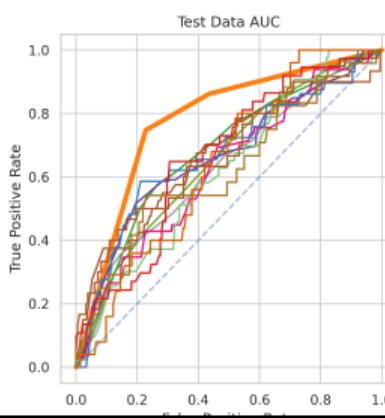
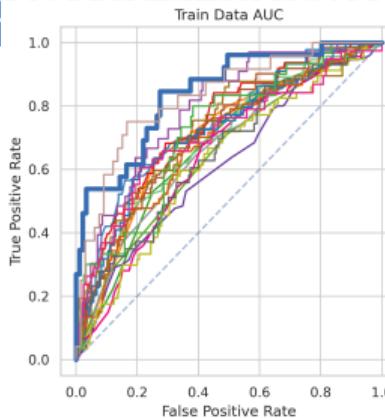
$$\mathbf{Y} = \mathbf{X}\vec{\beta}$$

β_0 is 0

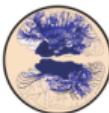
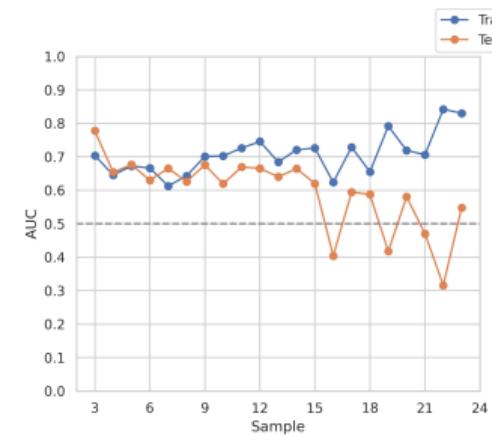


Hypothesis

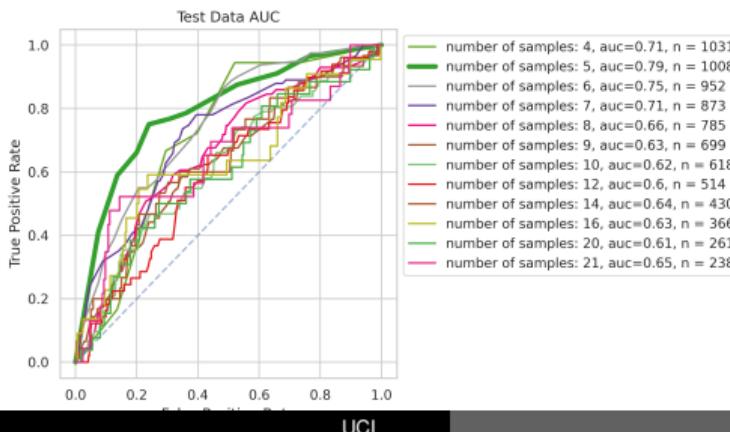
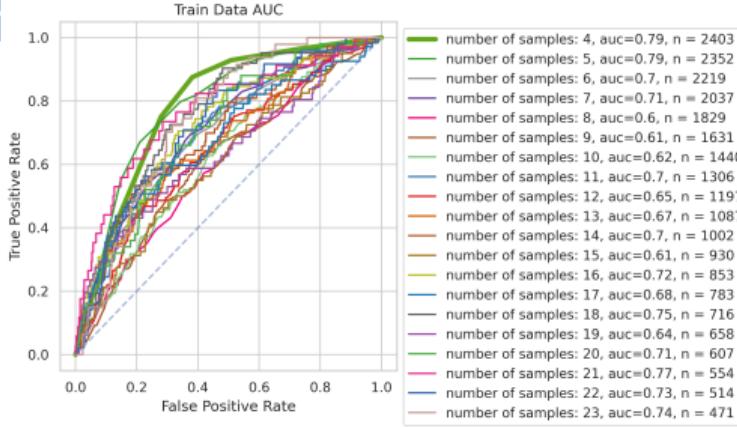
Walk - Logistic



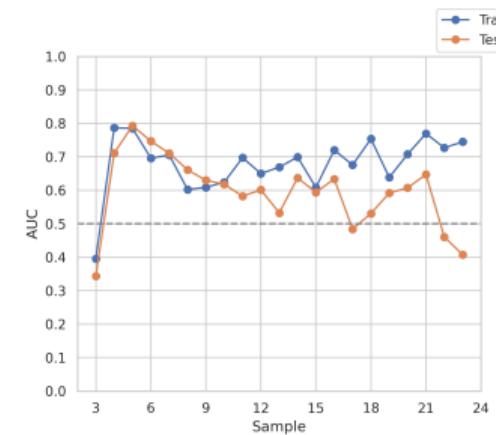
250ms



Hypothesis Testing - Weighted Sum of Random Walk - Logistic

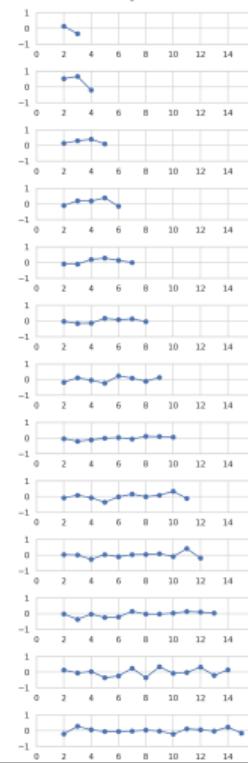
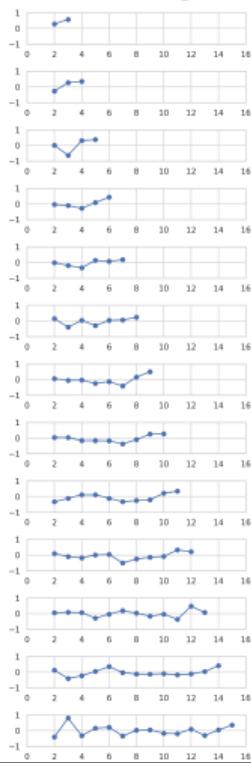


100ms



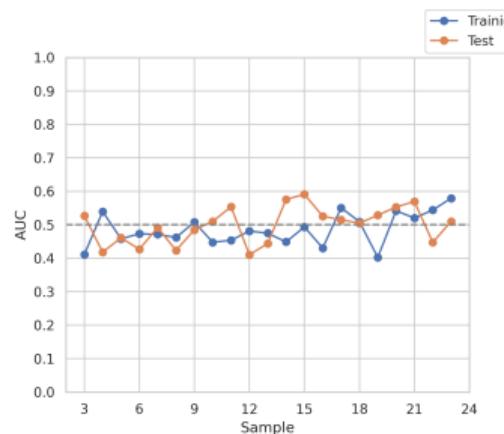
Hypothesis 3: Weighted Sum of Random Walk - Logistic

Visualizing the coefficients $\vec{\beta}$ for 250ms and 100ms ($n = 3, 4, \dots, 15$)

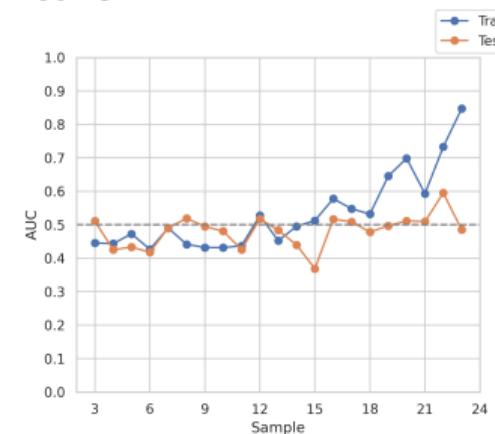


Hypothesis 5: Weighted Sum of Bernoulli - Logistic

performs poorly
100ms

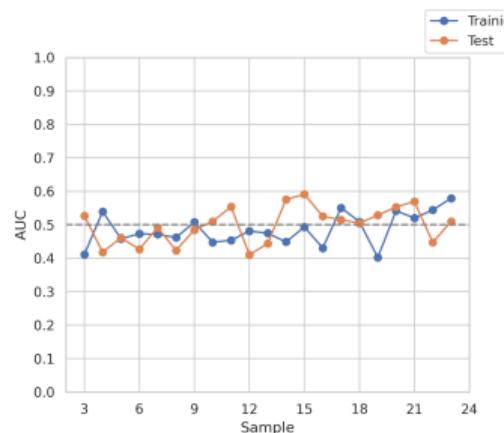


250ms

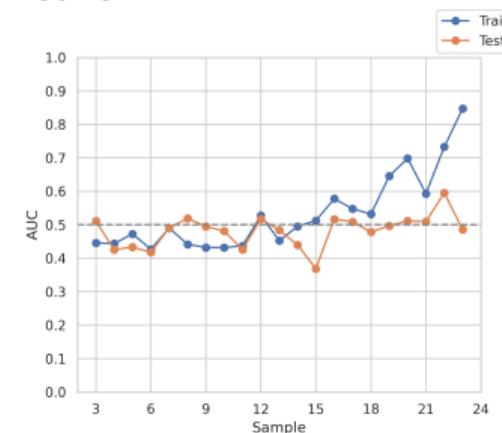


Hypothesis 4: Weighted Sum of Bernoulli - Logistic

performs poorly
100ms



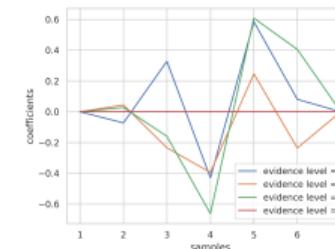
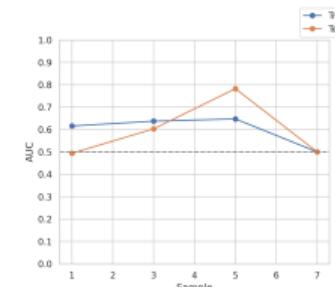
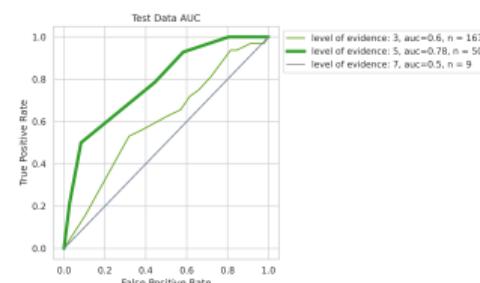
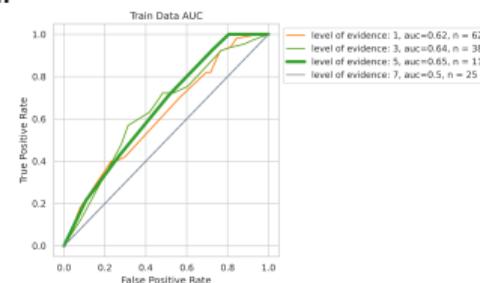
250ms



Hypothesis 5: Weighted Sum of Random Walk locked by evidence level - Logistic

at sample = 7, 250ms

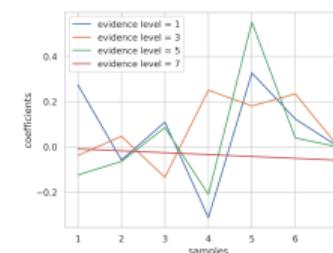
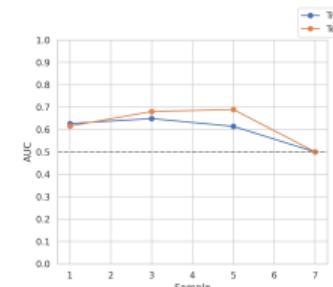
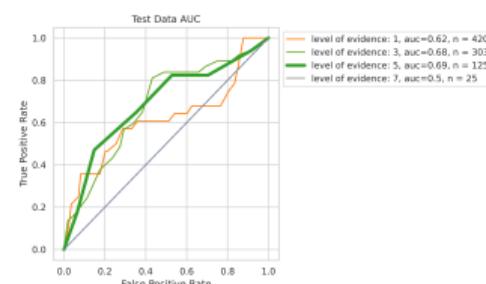
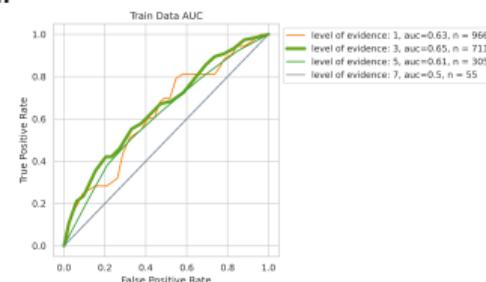
we find the possible level of evidence 1,3,5,7 at any point within the chain and back track 7 steps. zero pad if needed.



Hypothesis 5: Weighted Sum of Random Walk locked by evidence level - Logistic

at sample = 7, 100ms

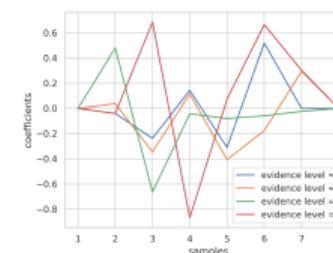
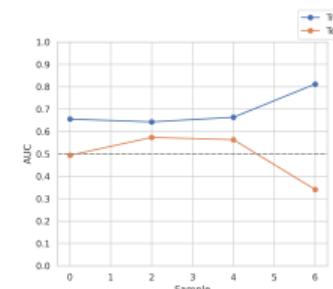
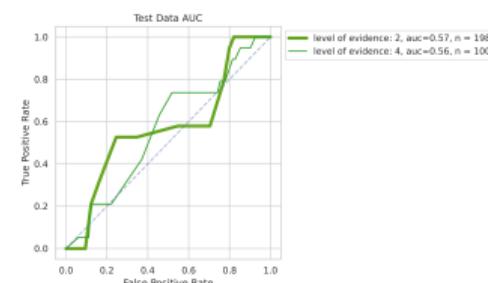
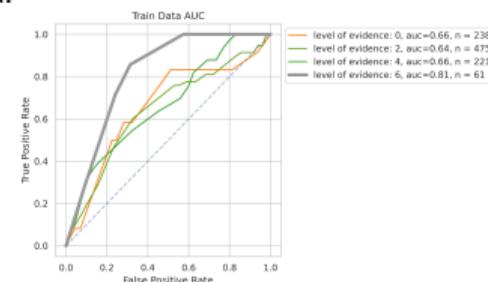
we find the possible level of evidence 1,3,5,7 at any point within the chain and back track 7 steps. zero pad if needed.



Hypothesis 5: Weighted Sum of Random Walk locked by evidence level - Logistic

at sample = 8, 250ms

we find the possible level of evidence 1,3,5,7 at any point within the chain and back track 7 steps. zero pad if needed.



Hypothesis 5: Weighted Sum of Random Walk locked by evidence level - Logistic

at sample = 8, 100ms

we find the possible level of evidence 1,3,5,7 at any point within the chain and back track 7 steps. zero pad if needed.

