Evidence Chain

University of California, Irvine

May 11, 2024

Sections

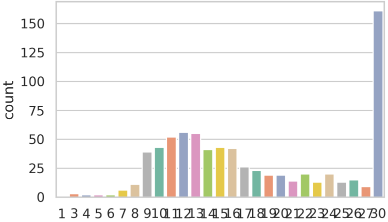
Dataset Characteristics

Research Question

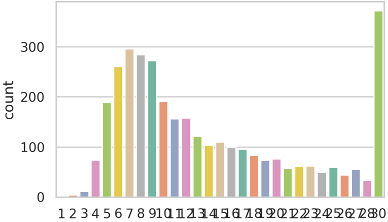
- Research questions: How do people accumulate evidence under the sequential sampling framework?
- Task: In the experiment, $+1$ means presenting one stimulus (stimulus "O"), and -1 means presenting the other stimulus (stimulus "X").
  On each trial, for $n \in N+$, let $\{X_n\}$ be a stochastic sequence 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

Number of Trials at Termination

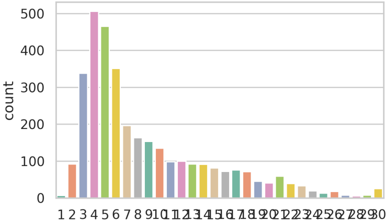
Sample Duration 0.05 s



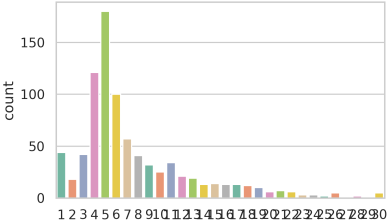
Sample Duration 0.1 s



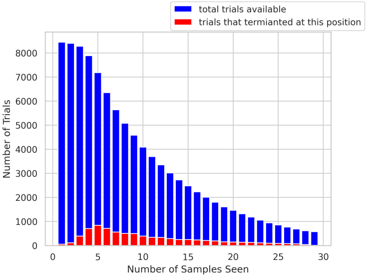
Sample Duration 0.25 s



Sample Duration 0.5 s



Percentage of Trials Terminated

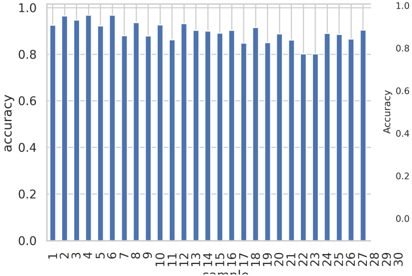


All conditions combined

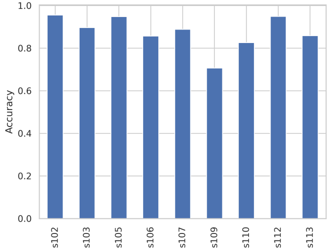
jenny TODO: 1: add termination by stimulus duration

Accuracy

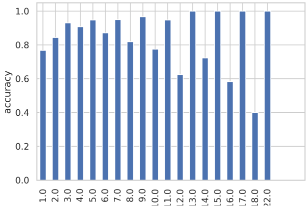
Accuracy by sample at termination



Accuracy by subject



Accuracy
by integrated bound at termination



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 $G_n : \{-1, 1\}, n \to \{0, 1\}$, such that:

$$1\{T = n\} = G_n\left(X_0, X_1, ..., X_n\right), \text{for all } n \in \mathbb{N}^+ \tag{1}$$

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

For all k trials $\vec{X}_1, \vec{X}_1 ... \vec{X}_k$, we find the trials where the stopping time $T >= n_0$, and classify the chains that terminated $G_n = 1$ versus the chains that proceeded $G_n = 1$.

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 topping criteria is deterministic, a fixed bound b_0 is used regardless number of samples.

$$G_n\left(X_0, X_1, ...X_n\right) = \begin{cases} 1, & S_n = b_0 \\ 0, & S_n \neq b_0 \end{cases} \tag{2}$$

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

$$G_n\left(X_0, X_1, ...X_n\right) = \begin{cases} 1, & S_n = f(S_n, n) \\ 0, & S_n \neq f(S_n, n) \end{cases} \tag{3}$$

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

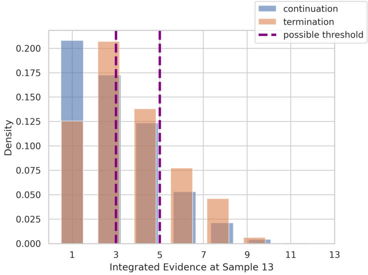
jenny TODO: 2: insert my proof next page

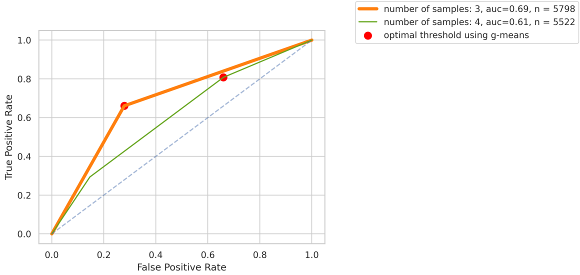
Hypothesis 1: Random Walk - fixed bound at all times vs. bound as a function of samples

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

- input = S_n at sample k where $k \in [3, 24]$ ism label = True if the chain is terminated; false if the chain continues

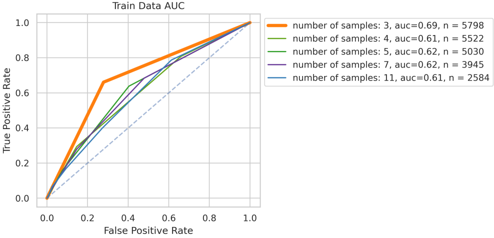
$$GSS = \sqrt{\text{sensitivity} * \text{specificity}} = \sqrt{(TPR * (1 - FPR)} \quad (4)$$

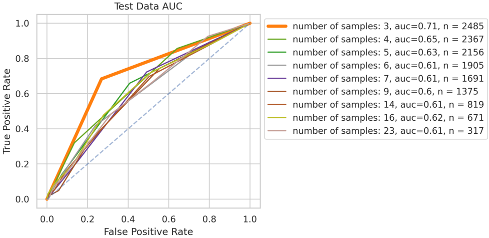


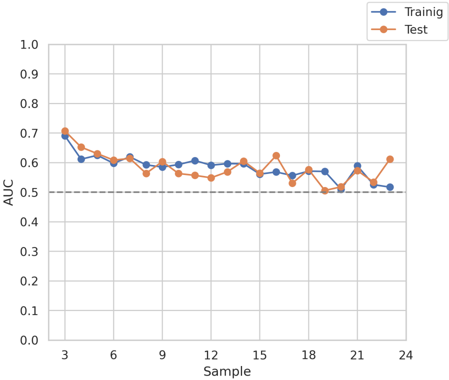


Hypothesis 1: Random Walk - fixed bound at all times vs. bound as a function of samples

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

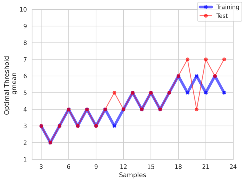


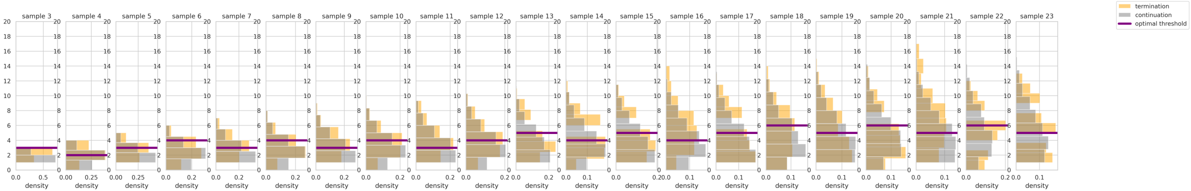




Hypothesis 1: Random Walk - fixed bound at all times vs. bound as a function of samples

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

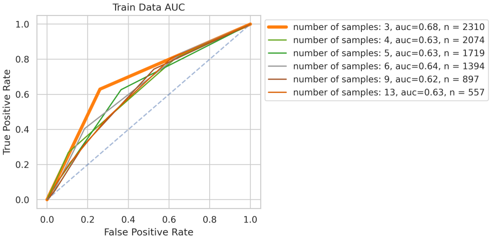


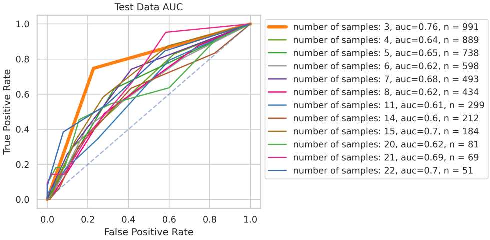


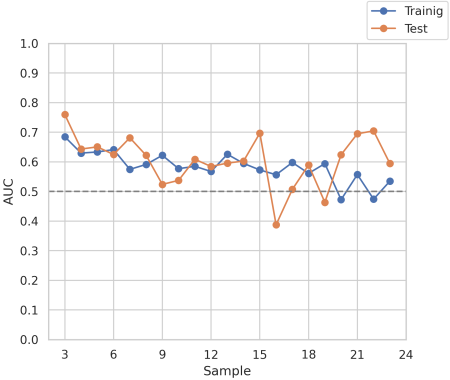
jenny TODO: 3: fix hist to be side to side

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

250ms

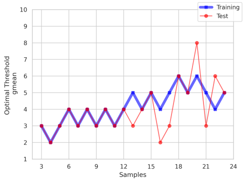


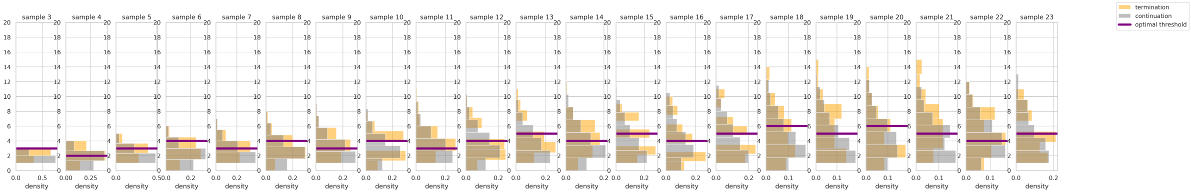




Hypothesis 1: Random Walk - fixed bound at all times vs. bound as a function of samples

250ms

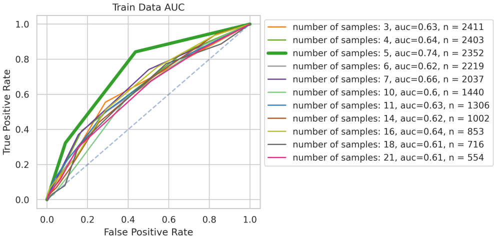


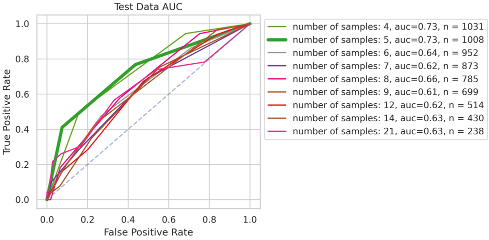


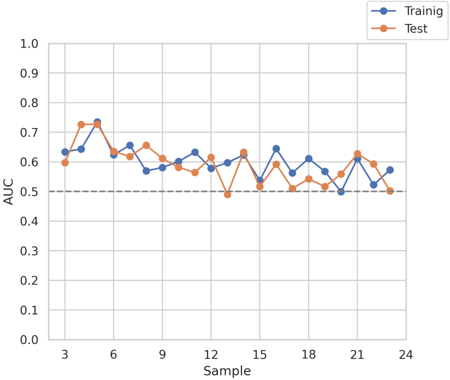
jenny TODO: 4: add this plot for 250ms and 100ms

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

100ms

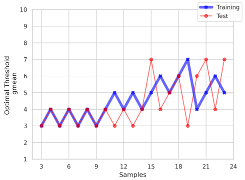


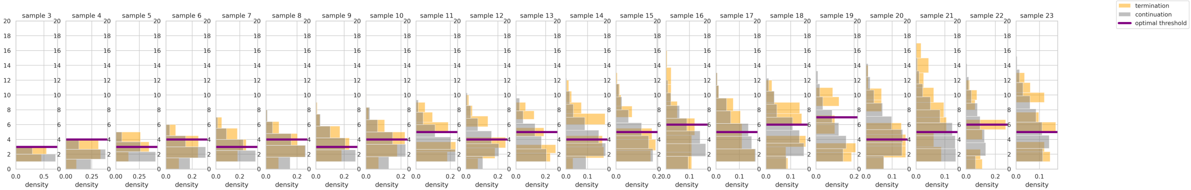




Hypothesis 1: Random Walk - fixed bound at all times vs. bound as a function of samples

100ms

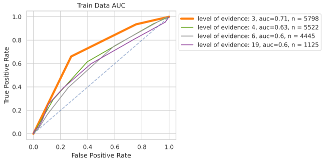


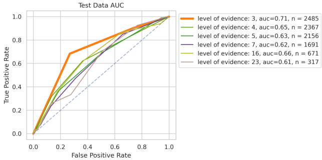


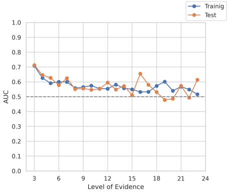
jenny TODO: 5: add this plot for 250ms and 100ms

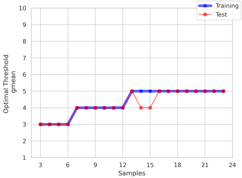
Hypothesis 2: Max Runs Model- fixed runs at all times vs. runs as a function of samples

Combined





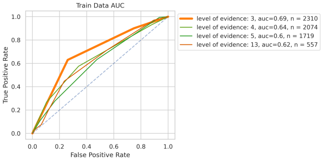


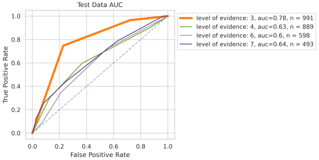


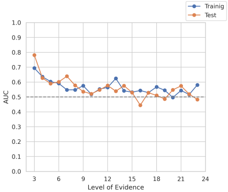
jenny TODO: 6: include a formula here

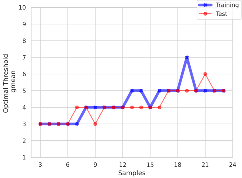
Hypothesis 2: Max Runs Model– fixed runs at all times vs. runs as a function of samples

250ms



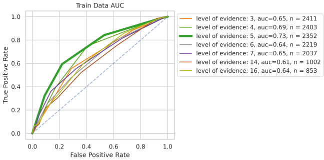


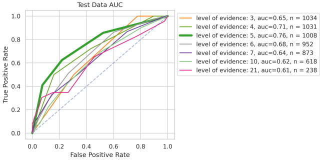


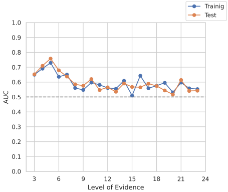


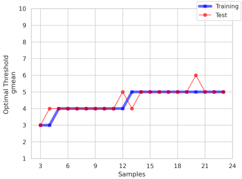
Hypothesis 2: Max Runs Model– fixed runs at all times vs. runs as a function of samples

100ms









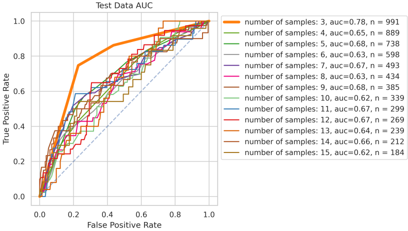
Hypothesis 3: Weighted Sum of Random Walk - Logistic

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

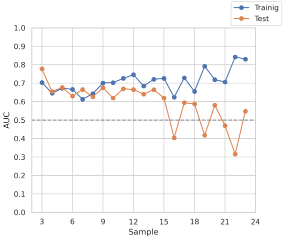
We will fit the model without intercept ($\beta_0 = 0$) as it is fair to assume the probability of termination is 0.5.

We will only look at chains with the same stimulus duration to interpret the coefficients.

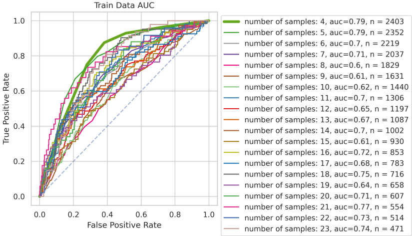
Hypothesis 3: Weighted Sum of Random Walk - Logistic

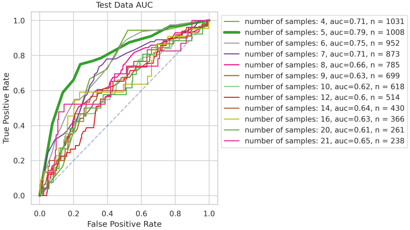


250ms

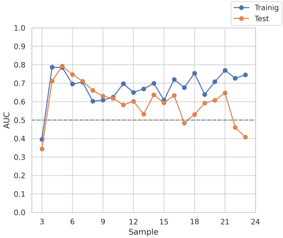


Hypothesis 3: Weighted Sum of Random Walk - Logistic



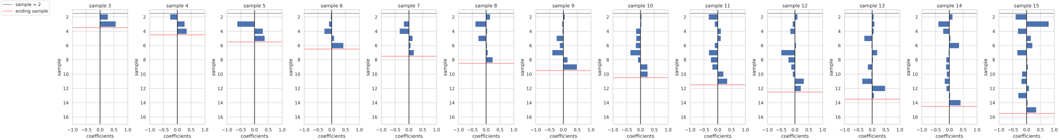


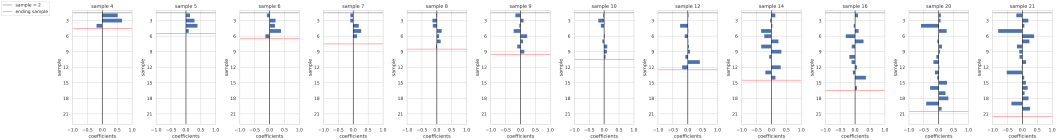
100ms



Hypothesis 3: Weighted Sum of Random Walk - Logistic Regression Coefficients

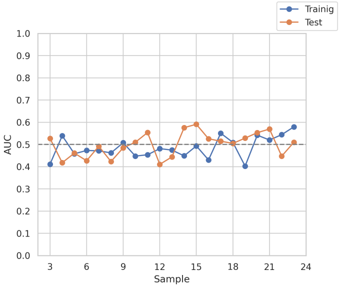
Visualizing the coefficients $\vec{\beta}$ for 250ms (top) and 100ms (bottom) ($n = 3, 4, ..., 15$)
We picked the models where test AUC is > 0.6



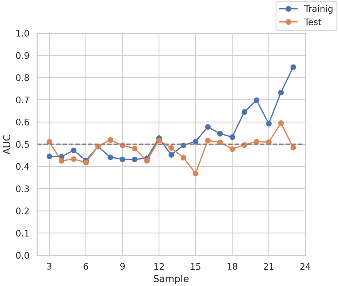


Hypothesis 4: Weighted Sum of Bernoulli - Logistic

Both performs poorly
100ms



250ms

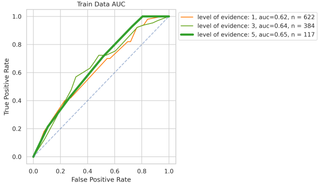


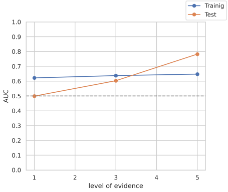
Hypothesis 3: Weighted Sum of Random Walk Same Level Evidence Level - Logistic

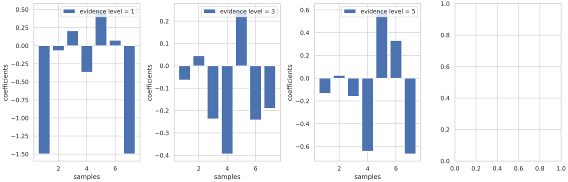
- Weighted Sum of Random Walk is the strongest model so far, indicating that people are integrating evidence, and giving weights to recent samples
- Question: what makes people decide to weight the samples more to favor the odds of stopping? Given the same level of evidence but different structures, do people still weight the recent samples

Hypothesis 3: Weighted Sum of Random Walk Same Level 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.

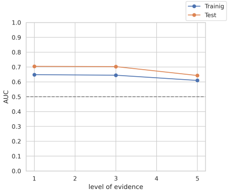


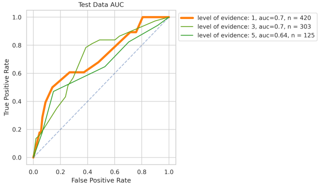


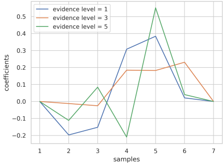


Hypothesis 3: Weighted Sum of Random Walk locked by evidence level - Logistic

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 by level of evidence at fixed length - Logistic

at sample = 8, 100ms
we find the possible level of evidence 2,4,6,8