**Introduction**

This project investigates the robustness of the Drift Diffusion Model (DDM) when some trials in the dataset are not generated from the DDM process but instead represent random guessing due to lapses in attention, mind-wandering, or distraction. While the DDM is widely used to model binary decision-making with accompanying reaction times, it typically assumes that all trials are the result of a noisy evidence accumulation process. In practice, this assumption may be violated when participants respond randomly or inattentively on some trials. This project explores how such contaminated trials affect parameter recovery and inference quality.

**Methods**

**Model Formulation and Assumptions**

We use the standard Drift Diffusion Model to generate decision-making behavior on binary choice tasks. The DDM assumes that noisy evidence accumulates over time until it reaches one of two decision boundaries. A decision is made once a boundary is hit. Key assumptions include:

* Evidence accumulation is continuous and stochastic.
* Decisions are binary.
* Non-decision time (such as perceptual and motor delays) is additive and fixed per trial.

In this project, a subset of trials (for example, 0 percent, 10 percent, or 20 percent) are deliberately contaminated by injecting data that does not follow the DDM process. These trials consist of randomly selected choices and uniformly sampled reaction times.

**Key Model Parameters**

The parameters of interest are:

* Drift rate (v): the average rate of evidence accumulation.
* Boundary separation (a): the distance between decision thresholds.
* Non-decision time (t0): a fixed delay unrelated to decision-making.

In this project, noise (sigma) is held constant to avoid identifiability issues. Instead, we introduce trial-level noise by adding non-DDM trials. The goal is to determine whether the core DDM parameters can still be reliably estimated in the presence of such contamination.

**Priors and Full Model Specification**

Although full Bayesian inference is not implemented in this version of the code, we conceptually assume standard weakly informative priors:

* v is normally distributed with mean 0 and standard deviation 1.
* a follows a half-normal distribution centered at 1.
* t0 is uniformly distributed between 0.2 and 0.5 seconds.

The likelihood assumes a diffusion process for clean DDM trials and a uniform distribution for noise trials.

**Data and Simulation Setup**

Each simulation condition includes 500 trials. Contamination levels are set to 0 percent, 10 percent, and 20 percent. For each trial:

* DDM trials are simulated using evidence accumulation until a boundary is crossed.
* Noise trials use a random choice (either 0 or 1) and a reaction time sampled from a uniform distribution between 0.3 and 2.5 seconds.

Each trial is labeled as either "ddm" or "noise." Multiple simulations are performed for each setting to support reliable conclusions.

**Software and Libraries**

The project uses Python 3.10 along with NumPy, Pandas, Matplotlib, and Seaborn.

**Evaluation Metrics**

The results are evaluated by comparing:

* Reaction time distributions using violin plots
* Choice proportions using bar plots
* Observed effects across clean versus contaminated conditions

This setup supports future extensions including full Bayesian estimation and parameter recovery metrics.

**Results**

**Diagnostics**

No formal sampling diagnostics are included since no MCMC estimation is run. However, simulations executed successfully and data generation was consistent across runs. The structure supports adding Bayesian inference in future stages.

**Interpretation**

The presence of noise trials affects both reaction time and choice distributions. As noise increases, reaction time distributions become broader and more uniform. The proportion of choice equal to 1 becomes closer to 50 percent, indicating random guessing. In contrast, clean trials show tighter reaction time clustering and more decisive choice patterns, aligned with the drift direction.

**Validation**

Ground truth parameters are known from simulation, which allows for qualitative validation. Clean conditions show clear decision dynamics, while noisy conditions flatten distributions and reduce observable effects of the model parameters.

**Sensitivity**

The results show that DDM inference is sensitive to the inclusion of contaminated trials. Even small levels of noise (such as 20 percent) significantly degrade the clarity of the data. This makes parameter recovery more difficult and highlights the importance of identifying or modeling such noise during analysis.

**Discussion**

This project demonstrates that the Drift Diffusion Model, while effective for modeling clean decision data, can struggle in the presence of even a modest number of random guessing trials. These trials disrupt the model's core assumptions and distort parameter inference. Future work should include:

* Implementing full Bayesian parameter recovery
* Applying posterior predictive checks
* Extending the DDM to include a mixture component that accounts for noise trials

This project emphasizes the need for robustness checks and preprocessing in cognitive modeling, especially in experiments where attention lapses or guessing may be present.