Replication Study : A generalized method for dynamic noise inference in modeling sequential decision-making

Süleyman Melih Portakal
Department of Computer Engineering, Boğaziçi University

Necdet Can Uzduran
Department of Computer Engineering, Boğaziçi University

Melih Gezer
Department of Computer Engineering, Boğaziçi University

URL for this document is :
https://www.overleaf.com/read/hbybsqrhgpcp#f00eaf

URL for our code repository :
https://github.com/canuzdrn/cognitive-science-final

Abstract

Introduction: Computational cognitive modeling is crucial for understanding decision making processes of humans and animals. Traditional models often assume static noise levels in choice behavior, which may lead to biased parameter estimates and model identification errors. This study proposes a method to dynamically infer noise, addressing the limitations of static noise assumptions.

Objective: To introduce and evaluate a dynamic noise inference method that models decision noise as varying over time, governed by transitions between discrete latent states (e.g., noisy states).

Methods: The proposed dynamic noise inference method allows noise levels to fluctuate based on a hidden Markov process that transitions between two latent states. The model's efficacy was tested using four empirical datasets involving both human and animal subjects performing various sequential decision making tasks. Model performance was assessed using maximum likelihood estimation and hierarchical Bayesian modeling.

Results: The dynamic noise inference method demonstrated improved model fit across all datasets compared to the static noise method. Notably, the method provided significant individual-level fit improvements for a substantial number of subjects. The dynamic method effectively captured periods of high noise and maintained the overall robustness of parameter recovery.

Conclusion: Dynamic noise inference enhances the accuracy of computational models in sequential decision-making tasks by accommodating temporal fluctuations in noise levels. This method is versatile, easily applicable into existing modeling frameworks, and beneficial for datasets with variable attention and noise patterns.

Keywords: add your choice of indexing terms or keywords; kindly use a semicolon; between each term

Information About the Original Study

Research Questions

The original study aimed to address the following research questions:

1. Can a dynamic noise inference method improve the fit of computational models in sequential decision making tasks compared to traditional static noise inference methods?

2. How do the dynamic and static noise inference methods compare in terms of model fit across different species, populations, tasks, and models?

3. Can the dynamic noise inference method accurately recover model parameters and reflect the true latent state transition probabilities in decision-making behavior?

Participants and Datasets

The original study utilized multiple datasets comprising both human and animal participants:

1. **Dynamic Foraging Dataset:** This dataset included choice behavior data from 48 mice performing a two-armed bandit task with changing reward probabilities.

2. **RLWM Dataset:** This dataset contained behavioral data from 91 human participants engaging in a task designed to test the interaction between **reinforcement learning and working memory.**

3. **Two-Step Tasks:** Two additional datasets included human behavioral data from adults and developmental populations performing different versions of the 2-step task, which differentiates between model based and model free reinforcement learning.

Design

The study employed a comparative design to evaluate the effectiveness of dynamic noise inference against static noise inference. In the context of static noise inference, noise is modeled as a constant parameter where decision maker is exposed to fixed level of disturbance. On the other hand, in dynamic noise inference, agent is transitioning between two latent states namely attentive state and noisy state. These latent states are design as Hidden Markov Processes that allow noise to vary over time.

Artifacts

1. Dynamic Foraging Dataset

2. RLWM (Reinforcement Learning and Working Memory) Dataset

Methodology

- **Model Evaluation**
  - Metrics: Akaike Information Criterion (AIC), one-tailed Wilcoxon signed-rank tests, and protected exceedance probability (pxp).
  - Simulations: Validation by simulating choice behavior and comparing to empirical data, with parameter and prediction probability recovery.
- **Statistical Analysis**
  - Software: Appropriate statistical tools for robust and valid analyses.

Summary of Results

The study found that the dynamic noise inference method generally improved model fit compared to the static noise inference method across all datasets. Key findings can be listed as such:

1. **Dynamic Foraging Dataset:** The dynamic noise model significantly outperformed the static model, especially for individual mice.
2. **RLWM Dataset:** The dynamic model showed better fit on average, although not significantly at the group level.
3. **Two-Step Tasks:** Similar patterns were observed, with dynamic models generally showing better fit but not significantly.

Overall, the dynamic noise inference method proved to be robust, accurately recovering model parameters and improving the interpretation of decision making behavior in various contexts. The method was shown to be versatile and easily integrated into existing model fitting procedures such as maximum likelihood estimation and hierarchical Bayesian modeling.

Information about Replication

Motivation for Conducting the Replication

There are several reasons for carrying out this study's replication. The primary goal of the replication was to confirm the findings of the first investigation and establish that the suggested dynamic noise inference approach systematically enhances model fit for a variety of datasets and circumstances. We hope to evaluate the robustness and generalisability of the results by repeating the study, especially with regard to how well the methodology adjusts to changes in participant populations (e.g., different species and demographic groups) and experimental tasks. Furthermore, by testing the dynamic noise inference approach with new or modified artifacts, the replication aims to extend the original results' reach and possibly uncover subtleties in its implementation and performance. This comprehensive examination not only reinforces the credibility of the original research but also contributes to the broader understanding of noise dynamics in decision-making processes.

Level of Interaction with the Original Experimenters

There was no communication with the original experimenters directly during this replication study. We, as Boğaziçi University Department of Computer Engineering students, carried out the replication on our own, using only the data from the published paper. The original researchers, Jing-Jing Li, Chengchun Shi, Lexin Li, and Anne G.E. Collins, were not consulted or offered any direction. Furthermore, we did not make use of any supplemental materials or published laboratory packages that might have been connected to the initial investigation. We only used the procedures and explanations provided in the original publication as the foundation for our replication efforts. This methodology guarantees the autonomous derivation of our findings and underscores the reproducibility and lucidity of the initial investigation.

Changes to the Original Experiment

In order to understand and apply the algorithms, we replicated the original study with sample dataset. Five features that were semantically similar to those used in the original study made up this dataset. Using this mock dataset, the relevant functions from the original paper were implemented, allowing the model to be used for both static and dynamic noise inference. This change was made to allow us to fully understand the algorithmic processes and to ensure that our implementation was accurate before applying it to the actual datasets. To further validate our results, we made use of the original RLWM dataset that the original researchers had provided. By utilizing both the mock and RLWM datasets, we were able to assess how well the dynamic noise inference method performed in various scenarios. Model fitting performance metrics, such as maximum likelihood estimation and the Akaike Information Criterion (AIC) score, were evaluated as part of our analysis. The necessity to verify that our procedures were correctly applied and comparable to those employed in the initial study, as well as to guarantee the robustness of our replication, served as the driving force behind these modifications.

Comparison of Replication Results to Original Results

Replication Study on Sample/Mock Dataset

Since the main purpose of this part of the study is to understand the logic and application of provided features and functions such as noise inference and AIC score calculation, we did not expect results comparable to the studies on the original datasets. Additionally, our sample dataset is too simple to capture the intricacies of real-world data. However, you can observe the findings of our replication study on the sample dataset below:

Replication Study on RLWM Dataset

As discussed in the original paper, we implemented an improved model based on later work instead of the model in

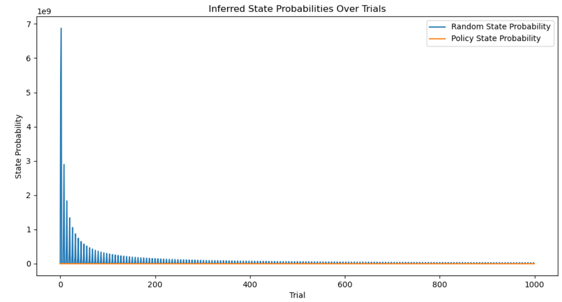

Figure 1: Plot that demonstrates the state that decision maker in on different trials (policy state vs random state)

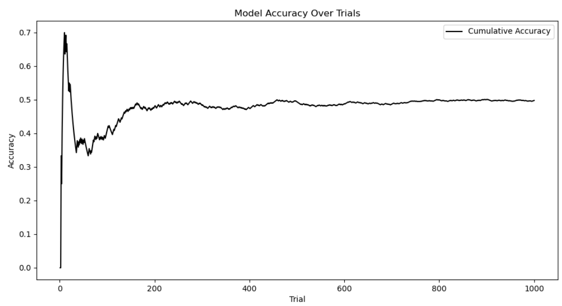

Figure 2: Accuracy (agent stayed on the policy state) vs trial plot

Collins' work (Master et al., 2020). We implemented this model using both static noise inference and dynamic noise inference. Then we compared the results of these models. As shown in Figure 4, we obtained better AIC scores using dynamic noise inference at both the group and individual levels. While the original paper reported no significant difference between dynamic and static noise inference at group level, our results indicate a significant difference. Potential reason of this could be differences in the implementation.

Conclusions Across Studies

Our results show that dynamic noise inference can improve model fit across diverse models. Dynamic noise inference is effective, versatile, and easy to be incorporated into existing model fitting procedures such as maximum likelihood estimation like we implemented in this replication study. We compared model fit using the AIC metric to show that dynamic noise inference improves fit at the group level and participant level.

In our replication study, we obtained slightly different figures compared to the original paper. The reasons for this could be that we do not know the exact implementation of the original model where it was trained in Matlab while we trained it in Python, and we might have obtained different re-

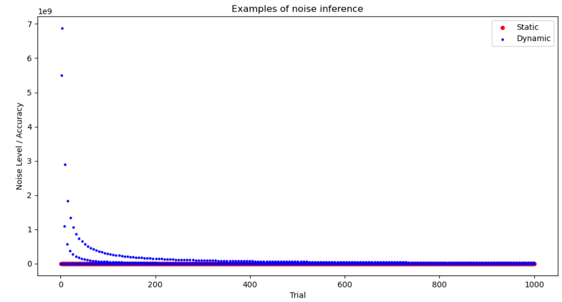

Figure 3: Amount of noise in different trials where we can observe the varying noise on dynamic noise inference and constant noise in static noise inference

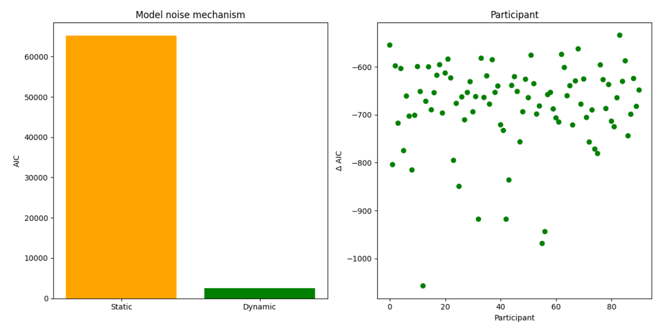

Figure 4: The differences in AIC between models with static and dynamic noise inference at the group (left) and individual (right) levels.

sults due to incorrect preprocessing techniques we've applied.

In summary, the dynamic noise inference method offers potential improvements over the static noise inference method currently used in decision-making behavior models. Our replication study and original study suggests that dynamic noise inference will enhance modeling in many decision-making scenarios, while keeping the model simple and making minimal assumptions.

Threats to Validities

Threats to validity is a term that is being used throughout research projects and replication studies to describe potential factors that can undermine the credibility or trustworthiness of the findings. These threats can affect whether the results of a study genuinely reflect reality or are influenced by other factors. Understanding and addressing these threats is crucial to ensure the robustness and generalization ability of research or study findings.

Threats to Internal Validity

Internal validity is the extent to which a research study establishes a trustworthy cause-and-effect relationship. This type of validity depends largely on the study's procedures and how rigorously it is performed.[1]

- **Confounding Variables** : In the context of the original study, potential confounding variables might include individual differences in attention, prior experience with similar tasks, and variability in task difficulty.

- **Bias during selection** : Ensuring random assignment of participants to different conditions is crucial. Any systematic differences between groups could lead to biased results.

- **Testing** : Repeated testing might affect participants' performance, so it's important to consider and mitigate these effects.

- **Tools and procedures** : Consistency in measurement tools and procedures is essential to avoid variability in data collection.

How to ensure internal validity ?

- **Randomization process** : Use random assignment to groups to control for confounding variables.

- **Blinding** : Implementing blind procedures to reduce bias from participants or researchers.

- **Standardization** : Standardize data collection procedures and use the same measurement instruments as the original study.

- **Study checks** : Conduct tests to identify potential issues in the study design or procedures.

Threats to External Validity

External validity is the validity of applying the conclusions of a scientific study outside the context of that study. In other words, it is the extent to which the results of a study can generalize or transport to other situations, people, stimuli, and times[2]

- **Population validity** : The sample used in the study should be representative of the broader population to which the findings will be generalized.

- **Temporal validity** : Consider whether the findings are applicable across different times or historical periods.

- **Ecological validity** : The study setting should accurately reflect real world conditions where the results are expected to apply.

- **Interactions** : Be mindful of how the interaction between different variables might limit generalizability.

How to ensure external validity ?

- **Real world settings** : Conduct the study in settings that closely mimic real world environments where the findings would apply.

- **Representative sampling** : Use a diverse and representative sample to enhance generalizability.

- **Replication across contexts** : Replicate the study across different settings and populations to confirm the robustness of findings.

Threats to Construct Validity

Construct validity concerns how well a set of indicators represent or reflect a concept that is not directly measurable. Construct validation is the accumulation of evidence to support the interpretation of what a measure reflects.[3]

- **Bias during measurement** : Systematic errors in measurement could skew results, so it's crucial to use reliable and valid instruments.

- **Bias during experiments** : Control for the influence of expectations of researchers on the measurement and interpretation of data.

- **Well-definedness of operational constructs** : Ensure that the constructs are clearly defined and appropriately measured. In the original study, constructs such as attention, noise levels, and decision policies need clear definitions.

How to ensure construct validity ?

- **Clear definitions** : Provide precise operational definitions for each construct being measured.

- **Measurement consistency** : Apply the same measurement procedures and instruments as the original study to ensure comparability.

- **Validating the instruments** : Use validated measurement tools that have been shown to accurately capture the used constructs throughout the study.

References

C., F., & V., W. (2021). *Computation noise in human learning and decision making: origin, impact, function*. Current Opinion in Behavioral Sciences.

C. D. Grossman, e. a. (2022). *Sero-tonin neurons modulate learning rate through uncertaint*. Current Biology.

Collins, A. G. (2018). *The tortoise and the hare: Interactions between reinforcement learning and working memory*. Journal of cognitive neuroscience.

G., C. A., & J., F. M. (2012). *How much of rein-forcement learning is working memory, not reinforcementlearning? a behavioral, computational, and neurogeneticanalysis*. European Journal of Neuroscience.

[1] https://www.verywellmind.com/internal-and-external-validity-4584479

[2] https://en.wikipedia.org/wiki/External_validity

[3] https://en.wikipedia.org/wiki/Construct_validity

Li, J., & Collins, A. G. (2023). *A generalized method for dynamic noise inference in modeling sequential decision-making*. Proceedings of the Annual Meeting of the Cognitive Science Society.