

EM-Driven GMM Optimization for the Randall Dataset: Rooting and Recall

Prabhav Kalaghatgi¹

¹Independent Researcher, Hyderabad, India

August 25, 2025

Abstract

1 Introduction

Related work

Claims to clarify

2 Results

2.1 Impact of initial parameters on final log likelihood scores

[width=0.8]"/home/pk/projects/prabhavk.github.io/public/figures/wasm-1756002951415_violin_l_f_inal_violin_t20250824225734Z.png"

Figure 1: This is the caption describing the figure.

2.2 Distribution of optimization scores across root locations

[width=0.8]path/to/your/figure.png

Figure 2: This is the caption describing the figure.

2.3 Wilcoxon-Mann-Whitney tests for difference in log likelihood scores

2.3.1 Initialization Method

2.3.2 Change across root locations

2.4 Significance of recall values

3 Methods

3.1 EM for fixed topology and root

3.2 Parameter initialization

3.2.1 Dirichlet

3.2.2 Parsimony

3.2.3 SSH

3.3 Statistical tests

3.4 Reproducibility of results

4 Discussion