Modeling Paul's Principle

1 Model Overview

We implement a continuous-time Hidden Markov Model (HMM) to test Paul's Principle: verbs with higher irregularity (both vowel and consonant alternations) regularize slower than those with lower irregularity. The model has:

- **Hidden states**: 4 states representing presence/absence of alternations: (1,1), (1,0), (0,1), (0,0)
- **Time span**: Middle High German to Early New High German
- **Key hypothesis**: Transitions from (1,1) $\rightarrow$ other states occur slower than from (1,0)/(0,1) $\rightarrow$ (0,0)

2 Data Structure

- N_{verbs}: Number of verb forms
- N_{obs}: Total observations
- $N_{\text{states}} = 4$: Hidden states
- `form`: Verb form ID
- `time`: Observation time
- `obs_v`, `obs_c`: Observed vowel/consonant alternations (binary)
- `freq`: Verb frequency
- `dialect_id`: Dialect identifier
- `lemma_id`: Verb lemma identifier
- `principal_part_id`: Principal part ID (1=Infinitive, 2=Past 3SG, 3=Past Pl, 4=Past Participle)

3 Irregularity Index Calculation

Precomputed irregularity index measures paradigm diversity:

$$\text{Irregularity} = \alpha \cdot \frac{|V| - 1}{m - 1} + (1 - \alpha) \cdot \frac{|C| - 1}{m - 1}$$

- $|V|$, $|C|$: Distinct vowel/consonant patterns per lemma per time point
- m: Observed principal parts per lemma per time point
- $\alpha = 0.5$: Balancing weight. We can adjust this number if we want to weight there to be more contribution to irregularity coming from vowel or consonant. Equal weighting for now.

The index is forward-filled when data is missing. Stored in matrices:

- `irregularity_index`: $N_{\text{lemmas}} \times n_{\text{time\_points}}$
- `m_values`: Number of observed principal parts. Will be used for measurement error adjustment.

4 Parameters

Transition dynamics

`log_lambda` $\in R^{4\times4}$	(Baseline log transition rates)
`beta_trans` $\in R^{4\times4}$	(Frequency effect)
$\gamma_{\text{irreg}} \in R$	(Paul's Principle effect)

Emission model

$\beta_v^{\text{true1}}, \beta_v^{\text{true0}} \in R$	(Vowel emission)
$\beta_c^{\text{true1}}, \beta_c^{\text{true0}} \in R$	(Consonant emission)
$\beta_{\text{dialects}} \in R^{n_{\text{dialects}}}$	(Dialect effects)

Paradigm interactions

`paradigm_influence` $\in R^{4\times4}$	(Cross-part influence)
$\sigma_{\text{adj}} \in R^+$	(Adjustment SD)
$\delta_l \in R^+$	(Lemma-specific adjustments)

5 Model Components

5.1 State-Space Encoding

States represent alternation patterns:

State	Vowel	Consonant
1	1	1
2	1	0
3	0	1
4	0	0

5.2 Adjusted Irregularity Index

Accounts for missing principal parts:

$$\text{true\_irregularity}_{l,t} = \sigma\left(\text{irregularity\_index}_{l,t} + \delta_l \cdot \frac{N_{\text{pp}} - m_{l,t}}{N_{\text{pp}}}\right)$$

where σ is logistic function, $N_{\text{pp}} = 4$ (principal parts), and $\delta_l \sim \mathcal{N}^+(0, \sigma_{\text{adj}})$.

TODO: I realize there probably is an error and logistic function should only be applied to the adjustment part, not the whole one.

5.3 Transition Rate Matrix

Instantaneous rate Q_{ij} from state i to j ($i \neq j$):

$$Q_{ij} = \exp\left(\underbrace{\texttt{log\_lambda}_{ij}}_{\text{baseline}} + \underbrace{\texttt{beta\_trans}_{ij} \cdot \text{freq}}_{\text{frequency}} - \underbrace{\gamma_{\text{irreg}} \cdot I_{\text{reg}}(i,j) \cdot \text{true\_irregularity}}_{\text{Paul's effect}} + \underbrace{\text{influence}_j}_{\text{paradigm}}\right)$$

- $I_{\text{reg}}(i,j) = 1, 0$ otherwise if transition reduces alternations (regularization). Since Paul's principle does not imply anything about the process of irregularization, we consider there to be no effect whatsoever for these transitions.
- Diagonal: $Q_{ii} = -\sum_{j \neq i} Q_{ij}$
- It maybe makes sense to have a single log_lambda$_{ij}$ for all transitions? Then, all the difference in transitions would be produced due to other variables. On the other hand, we can still infer some intercept per transition type independent of other variables.

5.4 Paradigm Influence Mechanism

The model now incorporates cross-paradigm influences to account for intra-paradigmatic leveling. This is implemented through:

1. **Influence Matrix**: A parameter matrix $\Gamma \in R^{N_{\text{pp}} \times N_{\text{pp}}}$ where $\Gamma_{pq} =$ `paradigm_influence`$[p,q]$ quantifies directional influence from principal part q to part p
   - Diagonal elements $\Gamma_{pp} = 0$ (no self-influence)
   - Off-diagonal elements $\Gamma_{pq} \sim \mathcal{N}(0.5, 0.5)$. The influence is likely either going to be there, hence positive bias via prior, or going to be zero.

2. **State Probability Estimation**: For each observation of principal part p at time t:

$$\text{For each } q \neq p: \quad \mathbf{w}_q = [w_{q1}, w_{q2}, w_{q3}, w_{q4}]$$
$$w_{qs} = P(s_q = s \mid \mathbf{obs}_q) \propto P(\mathbf{obs}_q \mid s_q = s)$$

   where $P(\mathbf{obs}_q \mid s_q = s)$ is computed using emission model parameters

3. **Influence Aggregation**: The net influence toward state s for part p:

$$\text{Influence}_s = \sum_{q \neq p} \Gamma_{pq} \cdot w_{qs}$$

4. **Transition Rate Modulation**: The influence directly affects transition rates:

$$Q_{ij} = \exp\left(\cdots + \text{Influence}_j\right)$$

   Higher Influence_j increases transition rate to state j

This mechanism captures two key phenomena:

- **Paradigm Uniformity Pressure**: When multiple principal parts share similar irregular patterns (w_{qs} concentrated on particular states), they mutually reinforce maintenance of those patterns
- **Asymmetric Leveling**: Directional influences ($\Gamma_{pq} \neq \Gamma_{qp}$) allow modeling cases where the influence of one part on another may be asymmetric or non-existent for certain combinations of principal parts.

Example: For verb *verliesen*:

- If past tense 3Pl (*verluren*) shows strong irregularity ($\mathbf{w}_{\text{past pl}} \approx [1, 0, 0, 0]$)
- While past tense 3SG (*verlos*) shows medium irregularity ($\mathbf{w}_{\text{past sg}} \approx [0, 1, 0, 0]$)
- And $\Gamma_{\text{past pl,past sg}} > 0$
- Then past plural (*verluren*) experiences pressure towards more regular state (0, 1). That's assuming their emission probabilities are equal, as they weight contributions of these cross-paradigm effects.

5.5 Emission Probabilities

Observation likelihood given state s:

$$P(\texttt{obs\_v} = 1 \mid s) = \sigma\left(\beta_v^{\text{state}} + \beta_{\text{dialects}}[d]\right)$$
$$\beta_v^{\text{state}} = \begin{cases} \beta_v^{\text{true1}} & \text{if } s \in \{1, 2\} \\ \beta_v^{\text{true0}} & \text{otherwise} \end{cases}$$

(Similarly for consonants)

6 Prior Distributions

$$
\begin{aligned}
&\texttt{log\_lambda}, \texttt{beta\_trans} \sim \mathcal{N}(0,1) \\
&\beta_v^{\text{true}*}, \beta_c^{\text{true}*} \sim \mathcal{N}(0,1) \\
&\gamma_{\text{irreg}} \sim \mathcal{N}(0,1) \\
&\texttt{paradigm\_influence} \sim \mathcal{N}(0.5, 0.5) \\
&\sigma_{\text{adj}} \sim \text{Exponential}(2) \\
&\delta_l \sim \mathcal{N}^+(0, \sigma_{\text{adj}}) \\
&\text{Initial state} \sim \text{Dirichlet}([3,2,2,1])
\end{aligned}
$$

7 Likelihood Computation

Using forward algorithm for each verb:

Algorithm 1 Forward Algorithm

Initialize $\text{log\_forward}_1 = \log(\text{initial\_probs}) + \log P(\mathbf{obs}_1 \mid s)$
for $t = 2$ to T **do**
  Compute Q matrix at time t using current irregularity and paradigm influence
  $P_{\Delta t} = \exp(Q \cdot \Delta t)$
  $\text{log\_forward}_t = \log\left(\exp(\text{log\_forward}_{t-1}) \cdot P_{\Delta t}\right) + \log P(\mathbf{obs}_t \mid s)$
end for
Marginal likelihood $= \log \sum \exp(\text{log\_forward}_T)$

8 Testing Paul's Principle

The key test is the sign of γ_{irreg}:

- $\gamma_{\text{irreg}} > 0$: Supports Paul (irregularity slows regularization)
- $\gamma_{\text{irreg}} < 0$: Contradicts Paul

Directly output as `paradigm_effect` in generated quantities.