Normative growth curves for phoneme articulation

Analysis plan

Considerations

- **Measure**: max (or 75th) percentile of expected phoneme probability ✓
- **Timing of measurement:** Time of given percentile ✓
- **Model(s)**
- **Differentiation between phonemes**
- **Variation within vs. between**
- **Age-specific quantiles**
- **Threshold for `acquisition'**

Model

Outcome:

$$Y_{ijk} = \text{Max probability of expected phoneme } i \text{ for person } j \text{ on instance } k$$

Model: Beta regression, whereby Y_{ijk} is Beta-distributed with mean μ_{ij} and precision ϕ_{ij} for each instance k

Consideration. May need to aggregate outcome over instances. Easier to analyze aggregated outcome, but at the expense of interpretability.

Mean model

$$\mu_{ij} = \text{Logit}(a_i(\theta_{ij} - b_i)) = \frac{1}{1 + \text{Exp}(-a_i(\theta_{ij} - b_i))}$$

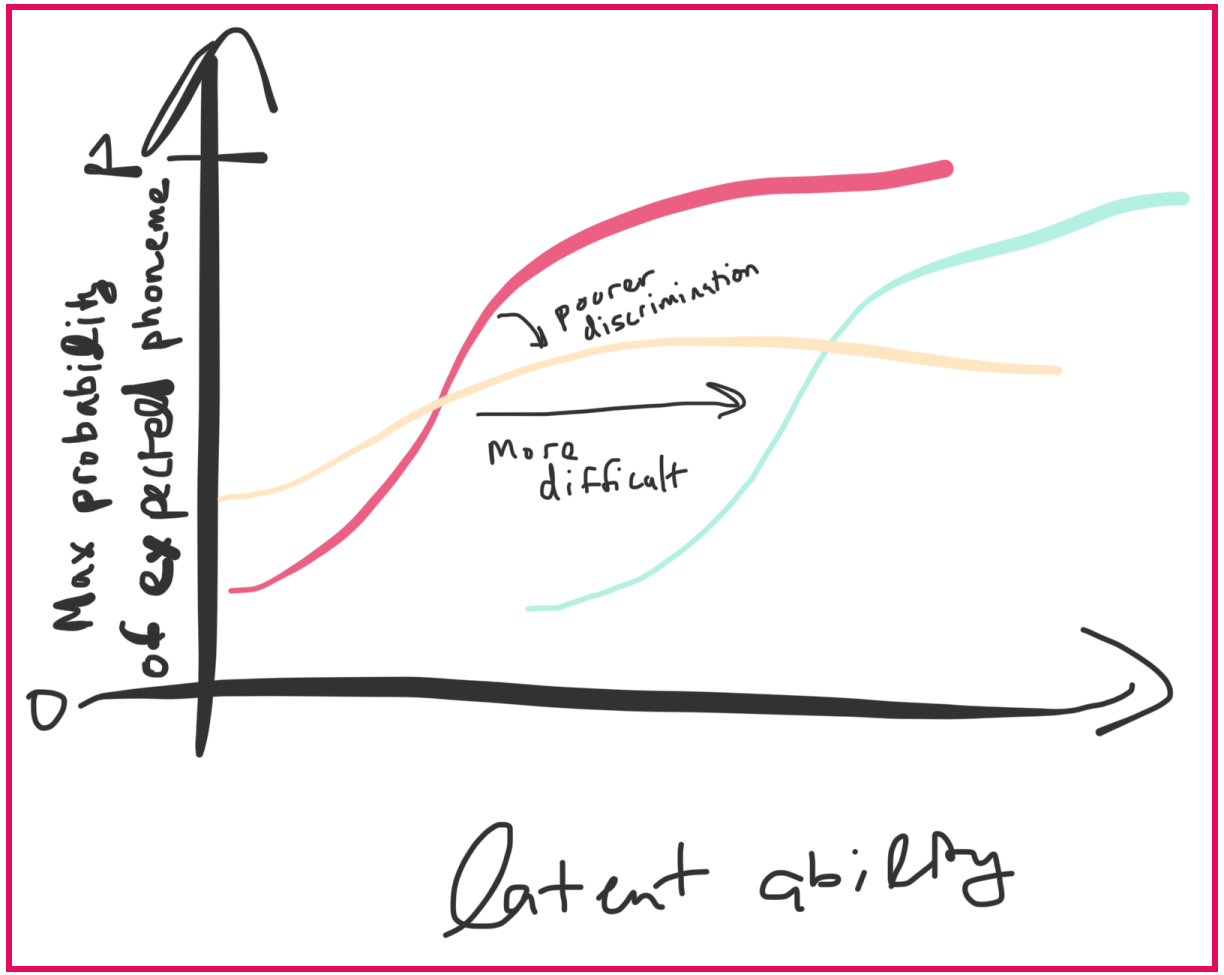
where

- a_i captures how strongly phoneme i **discriminates** between different levels of articulation quality
- b_i reflects the **difficulty** in articulating phoneme j
- θ_{ij} represents the **latent ability** of person i to articulate phoneme j

Consideration. May want to account for a ceiling effect on μ_{ij} for certain phonemes, meaning their probability never exceeds some value like 0.9.

Mean model

In a picture:



Latent ability

$$\theta_{ij} = \eta_{c_i j} + \beta_{c_i} \times \text{Age}_j$$

where

- c_i refers to a group of phonemes
- Age_j is age of person j
- $\eta_{c_i j}$ is latent ability of person i in articulating phenomes from group c_i
- β_{c_i} captures how much the latent ability in articulating phenomes from group c_i changes with age

Considerations. Will certainly need splines for age. May need to center age - if so, what is a good reference age? Age 0? Start with groupings driven by theory; later allow for data-driven groups

Precision model

Variance for Y_{ijk} is

$$\frac{\mu_{ij}(1-\mu_{ij})}{1+\phi_{ij}}$$

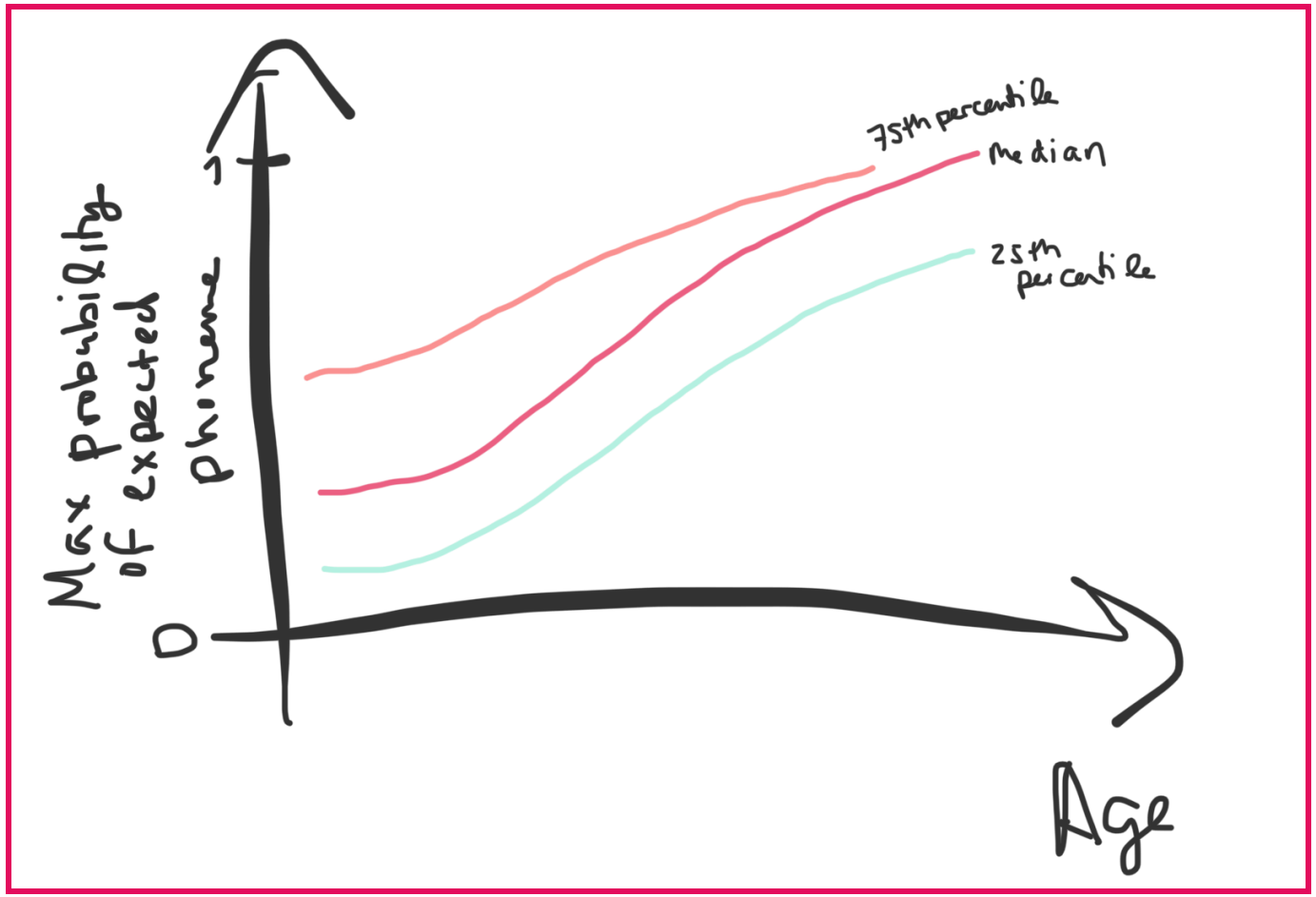
To start, assume $\phi_{ij} := \phi$ is constant.

What we can get from model

- **Age-specific percentiles.** For example:
  - The level of articulation for a given phoneme that 50% of TD children at age 4 are expected to surpass
  - The level of articulation for a given phoneme that 90% of TD children at age 6 are expected to surpass

What we can get from model

In a picture:



What we can get from model

- **Age's influence** on articulation for different phoneme groups:
  - Age when most TD children achieve a 90% level of articulation for a given phoneme group
  - Which groups of phonemes are mastered later in development
  - Age when greatest improvements in articulation are observed

What we can get from model

- **Discrimination** and **difficulty** parameters for each phoneme. For example:
  - Which phonemes are the most difficult to articulate
  - Which phonemes are best suited for discriminating the latent ability in articulation

What we can get from model

- How much variability in observed levels of articulation are attributable to differences in latent ability versus measurement noise