University of Antwerp
Faculty of Social Sciences

Objective rating method: Entropy

Speech intelligibility estimation

Jose Rivera

April 18, 2022

What are we going to talk about?

1. Preliminars

Research question

Research question

On two fronts:

1. Can comparative judgement (CJ) methods be used to assess speech intelligibility (SI)?,
   To investigate this wee need:
   - an objective measure of SI
2. where CJ stands versus absolute holistic judgement (HJ) methods?,
   In terms of:
   - validity
   - reliability
   - statistical efficiency
   - time efficiency

Objective measure of SI

the most objective (we know of) measure of SI comes from a transcription task:

1. transcribing children's utterances (made by multiple judges),
2. align transcriptions at the utterance level,
3. calculate an entropy measure (H), defined as

$$H = H(\mathbf{p}) = \frac{-\sum_{i=1}^{n} p_i \cdot \log_2(p_i)}{\log_2(N)}$$

4. characteristics of H [1, 3]
   - bounded in $[0, 1]$ space,
   - utterances with more agreement are more intelligible, and therefore $H \to 0$,
   - utterances with low agreement are less intelligible, and therefore $H \to 1$.

1. Preliminars

Research hypothesis production

A typical scientific lab[1]

What is needed?

1. Quality of theory
2. Quality of data
3. Reliable procedures and code
4. Quality of data analysis
5. Documentation
6. Reporting

What we will deal with:

1. Quality of theory
2. Quality of data
3. Reliable procedures and code
4. Quality of data analysis
5. Documentation
6. Reporting

[1] McElreath [7], lecture 20 and McElreath [8], chapter 17

Research hypothesis production[2]

Well known challenges

- Insufficient data
- Wrong population
- Measurement error
- Selection bias
- Confounding

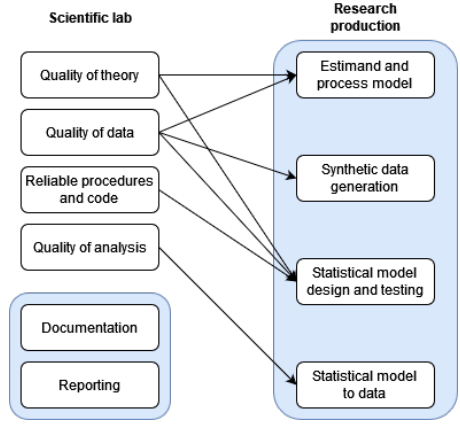
Known challenges in our research;

- Insufficient data (possibly)
- Wrong population
- Measurement error
- Selection bias
- Confounding

[2] Hernán [5], lesson 4

Research hypothesis schematics[3]

a. Estimand and process model
b. Synthetic data generation
c. Statistical model design and testing
d. Apply statistical model to data



[3]McElreath [8], lecture 20, Pearl [9]. Follow Fogarty et al. [4] on item (c).

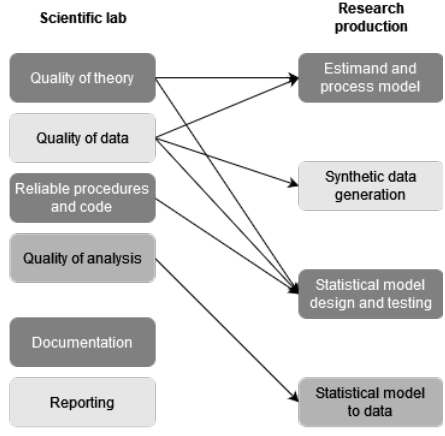
Why do we need to follow this?

Because the improvement of:

- A clear definition of the estimand and process model (assumptions).
- An improved the reliability of your procedures.
- As a documentation procedure.

leads to:

- A sound analysis, and sound results (even when we cannot answer our question).
- An improved planning to get data.

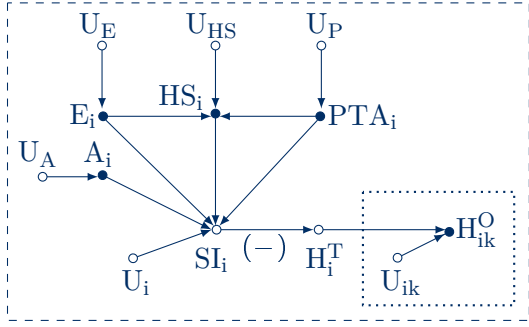


2. Research hypothesis procedure

Estimand and process model

The theory behind our research

- H_{ik} = (observed) entropy replicates
- H_i = (latent) child's entropy
- SI_i = (latent) child's SI score (inversely related to H_i^T)
- A_i = child's "hearing" age
- E_i = child's etiology of disease
- HS_i = child's hearing status
- PTA_i = child's pure tone average
- variables assumed independent, beyond the described relationships,

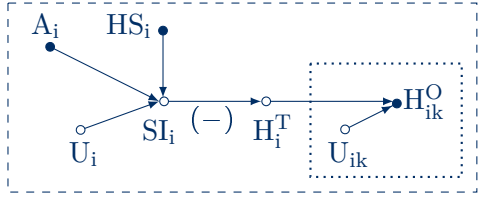


General causal diagram

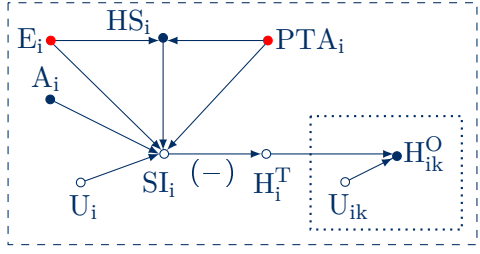
$$P(\mathbf{U}) = P(U_{ik}, U_i, U_A, U_E, U_{HS}, U_P)$$
$$= P(U_{ik})P(U_i)P(U_A)P(U_E)P(U_{HS})P(U_P)$$

Interested in two effects

1. total effects model inherits:
   - children's characteristics that lead to the fitting of specific apparatus,
   - the (convenience of) sample selection (fixed with post-stratification)
2. to do the last, we stratify for all variables that explain variability, ergo, use a direct effects model
3. two levels: replicates (k), children (i), denoted by discontinuous squares
4. U_{ik} = replicates measurement error
   U_i = between child SI variability



(b) total effects



(a) direct effects

Probabilistic (causal) model

First form

$$
\begin{aligned}
\mathrm{H}^{\mathrm{O}}_{\mathrm{ik}} &\sim \mathrm{BetapProp}(\mathrm{H}^{\mathrm{T}}_{\mathrm{i}}, \mathrm{M}_{\mathrm{ik}}) \\
\mathrm{H}^{\mathrm{T}}_{\mathrm{i}} &= \mathrm{inv\_logit}(-\mathrm{SI}_{\mathrm{i}}) \\
\mathrm{SI}_{\mathrm{i}} &\sim \mathrm{Normal}(\mu_{\mathrm{SI}}, \sigma_{\mathrm{Ui}}) \\
\mu_{\mathrm{SI}} &= \alpha + \alpha_{\mathrm{HS[i]}} + \alpha_{\mathrm{E[i]}} \\
&\quad + \beta_{\mathrm{A,HS[i]}}(\mathrm{A}_{\mathrm{i}} - \bar{\mathrm{A}}) + \beta_{\mathrm{P}}\mathrm{PTA}_{\mathrm{i}} \\
\mathrm{HS}_{\mathrm{i}} &\sim \mathrm{data} \\
\mathrm{A}_{\mathrm{i}} &\sim \mathrm{data} \\
\mathrm{E}_{\mathrm{i}} &\sim \mathrm{data} \\
\mathrm{PTA}_{\mathrm{i}} &\sim \mathrm{data} \\
\mathrm{U} &\sim \mathrm{unobservable}
\end{aligned}
$$

(a) general probabilistic model

$$
\begin{aligned}
\mathrm{H}^{\mathrm{O}}_{\mathrm{ik}} &\leftarrow \mathrm{f}(\mathrm{H}^{\mathrm{T}}_{\mathrm{i}}, \mathrm{U}_{\mathrm{ik}}) \\
\mathrm{H}^{\mathrm{T}}_{\mathrm{i}} &\leftarrow \mathrm{f}(\mathrm{SI}_{\mathrm{i}}) \\
\mathrm{SI}_{\mathrm{i}} &\leftarrow \mathrm{f}(\mathrm{HS}_{\mathrm{i}}, \mathrm{A}_{\mathrm{i}}, \mathrm{E}_{\mathrm{i}}, \mathrm{PTA}_{\mathrm{i}}, \mathrm{U}_{\mathrm{i}}) \\
\mathrm{HS}_{\mathrm{i}} &\leftarrow \mathrm{f}(\mathrm{U}_{\mathrm{HS}}) \\
\mathrm{A}_{\mathrm{i}} &\leftarrow \mathrm{f}(\mathrm{U}_{\mathrm{A}}) \\
\mathrm{E}_{\mathrm{i}} &\leftarrow \mathrm{f}(\mathrm{U}_{\mathrm{E}}) \\
\mathrm{PTA}_{\mathrm{i}} &\leftarrow \mathrm{f}(\mathrm{U}_{\mathrm{P}}) \\
\mathrm{U} &\sim \mathrm{P}(\mathbf{U})
\end{aligned}
$$

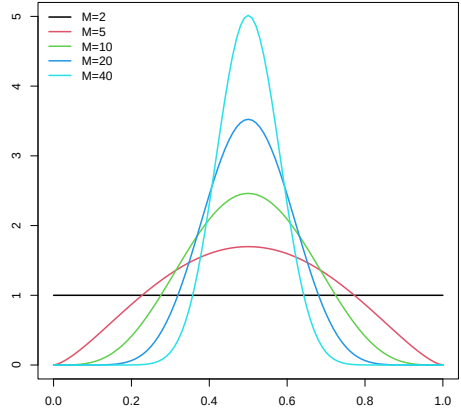
(a) general structural model

Probabilistic (causal) model

First form

$$
\begin{aligned}
\mathrm{H}^{\mathrm{O}}_{\mathrm{ik}} &\sim \mathrm{BetapProp}(\mathrm{H}^{\mathrm{T}}_{\mathrm{i}}, \mathrm{M}_{\mathrm{ik}}) \\
\mathrm{H}^{\mathrm{T}}_{\mathrm{i}} &= \mathrm{inv\_logit}(-\mathrm{SI}_{\mathrm{i}}) \\
\mathrm{SI}_{\mathrm{i}} &\sim \mathrm{Normal}(\mu_{\mathrm{SI}}, \sigma_{\mathrm{Ui}}) \\
\mu_{\mathrm{SI}} &= \alpha + \alpha_{\mathrm{HS[i]}} + \alpha_{\mathrm{E[i]}} \\
&\quad + \beta_{\mathrm{A,HS[i]}}(\mathrm{A}_{\mathrm{i}} - \bar{\mathrm{A}}) + \beta_{\mathrm{P}}\mathrm{PTA}_{\mathrm{i}} \\
\mathrm{HS}_{\mathrm{i}} &\sim \mathrm{data} \\
\mathrm{A}_{\mathrm{i}} &\sim \mathrm{data} \\
\mathrm{E}_{\mathrm{i}} &\sim \mathrm{data} \\
\mathrm{PTA}_{\mathrm{i}} &\sim \mathrm{data} \\
\mathrm{U} &\sim \mathrm{unobservable}
\end{aligned}
$$

(a) general probabilistic model



Probabilistic (causal) model

Second form

$$
\begin{aligned}
H^O_{ik} &\sim \text{BetapProp}(H^T_i, M_{ik}) \\
H^T_i &= \text{inv\_logit}(-SI_i) \\
SI_i &= a_i + \alpha + \alpha_{HS[i]} + \alpha_{E[i]} \\
&\quad + \beta_{A,HS[i]}(A_i - \bar{A}) + \beta_P PTA_i \\
a_i &\sim \text{Normal}(0, \sigma_{Ui}) \\
HS_i &\sim \text{data} \\
A_i &\sim \text{data} \\
E_i &\sim \text{data} \\
PTA_i &\sim \text{data} \\
U &\sim \text{unobservable}
\end{aligned}
$$

(a) general probabilistic model

$$
\begin{aligned}
H^O_{ik} &\leftarrow f(H^T_i, U_{ik}) \\
H^T_i &\leftarrow f(SI_i) \\
SI_i &\leftarrow f(HS_i, A_i, E_i, PTA_i, U_i) \\
HS_i &\leftarrow f(U_{HS}) \\
A_i &\leftarrow f(U_A) \\
E_i &\leftarrow f(U_E) \\
PTA_i &\leftarrow f(U_P) \\
U &\sim P(\mathbf{U})
\end{aligned}
$$

(a) general structural model

2. Research hypothesis procedure

Synthetic data generation

Idealized data[4]

Simulation data can serve as [6, 7],

1. A place where to test your model, on multiple purposes,
   - parameter recovery
   - power
2. A (possible) reflection of a population,
   - children's group proportion [2]
3. A (possible) reflection of a hypothesis,
   - size of effects (no previous information)

```
(sim_name=NULL, # file_name need to include
sim_save=NULL, # file_save need to include
seed=NULL, # seed
I=350, # experimental units (children)
K=10, # replicates (utterances)
p=c(0.50, 0.175, 0.325), # children prop. o
par=list( m_i=0, s_i=0.5, # hyperprior chil
          m_M=10, s_M=NULL, # generation of
          a=0, aE=-0.1, aHS=-0.4,
          bP=-0.1, bA=0.15, bAHS=0 ) ){
```

[4] more details in file: 1_2_E_sim_fun.R

Idealized data

About the size of the effects
(in logits, no previous info),

1. aE $= -0.1$, assumes E ordered by severity (it might not be possible),
2. aHS $= -0.4$, assumes 0.4 difference between NH children and HI/CI,
3. bP $= -0.1$, assumes decrease of SI per PTA unit
   ($+10$ PTA units $\Rightarrow -1$ logit),
4. bA $= -0.15$, assumes decrease of SI per A unit, beyond the minimum
   ($+10$ A units $\Rightarrow +1.5$ logits),

```
(sim_name=NULL, # file_name need to include
sim_save=NULL, # file_save need to include
seed=NULL, # seed
I=350, # experimental units (children)
K=10, # replicates (utterances)
p=c(0.50, 0.175, 0.325), # children prop. o
par=list( m_i=0, s_i=0.5, # hyperprior chil
          m_M=10, s_M=NULL, # generation of
          a=0, aE=-0.1, aHS=-0.4,
          bP=-0.1, bA=0.15, bAHS=0 ) ){
```

Idealized data

1. variables are generated in a random fashion
2. random effects define the between SI variability

```r
# 1. true data ####
dT = data.frame(matrix(NA, nrow=I, ncol=1))
names(dT) = c('child_id')
dT$child_id = 1:I


# assigning children to groups
n = round( p*I )
if( sum(n) != I ){
  if( I - sum(n[c(1,3)]) > n[2] ){
    n[2] = I - sum(n[c(1,3)]) # to sum the right amount
  } else {
    n[3] = I - sum(n[c(1,3)]) # to sum the right amount
  }
}


# generating covariates
if(!is.null(seed)){
  set.seed(seed+1)
}
dT$HS = c( rep(1, n[1]), rep(2, n[2]), rep(3, n[3]))
dT$A = round( rnorm( sum(n), 5, 1) )
dT$A = with(dT, ifelse(A>7, 7, A) )

dT$E = c( rep(1, n[1]), # no way to know true effects
          sample(2:3, size=n[2], replace=T),
          sample(3:4, size=n[3], replace=T))

dT$PTA = c( round(rnorm(n[1], 60, 15)), # first 12 NH
            round(rnorm(n[2], 90, 15)), # next 10
            round(rnorm(n[3], 110, 15))) # last 10


# children's random effects
if(!is.null(seed)){
  set.seed(seed-1)
}
par$re_i = rnorm(I, par$m_i, par$s_i)
dT$re_i = par$re_i # children's random effects (between SI v
```

Idealized data

1. we use second form of the probabilistic model
2. “true’ entropy (Ht) is inversely related to SI
3. we simulate measurement error through M from BetaProp() distribution (as previously shown).

```
# linear predictor / SI index
dT$SI = with(dT, re_i + par$a + par$aE*E + par$aHS*HS +
               par$bA*(A - min(A)) +
               par$bAHS*(A - min(A))*HS +
               par$bP * c( standardize(PTA) ) )

# true entropy
dT$Ht = inv_logit(-dT$SI) # true entropy (SI -> Ht: negative)


# variability of H
if(!is.null(seed)){
  set.seed(seed+2)
}
if( is.numeric(par$m_M) & !is.numeric(par$s_M) ){
  par$M = rep(par$m_M, I)
} else{
  par$M = round( rlnorm(I, meanlog=par$m_M, sdlog= par$s_M) )
}
dT$M = par$M # same df for all children (not same shape!!)


# rounding
dT[,6:ncol(dT)] = round( dT[,6:ncol(dT)], 5)



# 2. observed data ####
N = I*K
dO = data.frame(matrix(NA, nrow=N, ncol=3))
names(dO) = c('child_id','utt_id','H')
dO$child_id = rep(1:I, each=K)
dO$utt_id = rep(1:K, I)

# generating observed H
# i=1
if(!is.null(seed)){
  set.seed(seed-2)
}
```

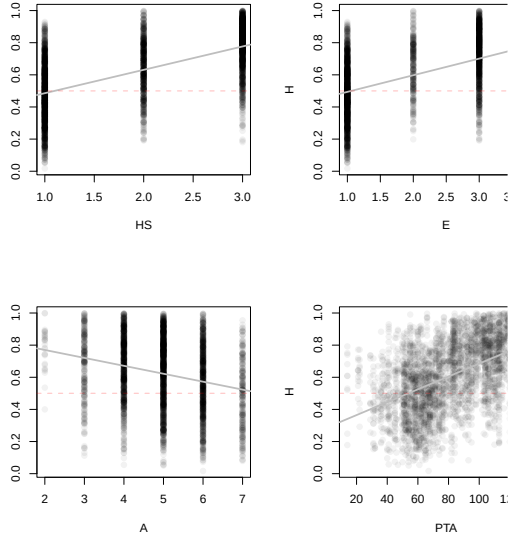
Idealized data

1. we simulate replicate measures of entropy (H)
2. we storage all relevant parameters and data

```
for(i in 1:I){

  # identify data
  idx = dO$child_id == i

  # linear predictor
  dO$H[idx] = rbeta2(n=K, prob=dT$Ht[i], theta=dT$M[i])
}

# round
dO$H = round( dO$H, 5)


# 3. list data ####
dL = list(
  # dimensions
  N = nrow(dO), # observations
  I = max(dO$child_id), # children
  K = max(dO$utt_id), # utterances

  # category numbers
  cHS = max(dT$HS),
  cE = max(dT$E),

  # child's data
  HS = dT$HS,
  Am = with(dT, A-min(A) ), # centered at minimum
  E = dT$E,
  sPTA = c( standardize( dT$PTA ) ),

  # observed data
  H = with(dO, ifelse(H==0, 0.0001, ifelse(H==1, 0.9999, H)) ),
  cid = dO$child_id,
  uid = dO$utt_id
)


# 4. save data ####
mom = list(dS=list( dT=dT, dO=dO, par=par), dL=dL)
```

Example

notice we can simulate any desired relationship, or lack of thereof.



2. Research hypothesis procedure

Statistical model design and testing

Model design[5]

Purpose:

- to have reliable procedures,
- to maintain a clear documentation,
- to have a sound analysis

```
transformed parameters{
    vector[I] SI;          // SI index (per child)
    vector[I] Ht;          // true entropy (per chi

    SI = a + re_i;         // linear predictor
    Ht = inv_logit(-SI);   // average entropy (SI -
}
model{

    // hyperpriors
    m_i ~ normal( 0 , 0.2 );
    s_i ~ exponential( 1 );

    // priors
    a ~ normal( 0 , 0.2 );
    re_i ~ normal( m_i , s_i );

    // likelihood
    for(n in 1:N){
      H[n] ~ beta_proportion( Ht[cid[n]] , 10 );
    }
}
```

[5] Following Fogarty et al. [4]

Model design

Procedure:

- step by step, instantiating one difficulty at the time
- Try the centered and non-centered versions

```
transformed parameters{
    vector[I] re_i;       // random intercepts (per
    vector[I] SI;         // SI index
    vector[I] Ht;         // true entropy (per child

    re_i = m_i + s_i*z_re;// non-centered RE
    SI = a + re_i;        // linear predictor
    Ht = inv_logit(-SI);  // average entropy (SI ->
}
model{

    // hyperpriors
    m_i ~ normal( 0 , 0.2 );
    s_i ~ exponential( 1 );

    // priors
    a ~ normal( 0 , 0.2 );
    z_re ~ std_normal();

    // likelihood
    for(n in 1:N){
      H[n] ~ beta_proportion( Ht[cid[n]] , 10 );
    }
}
```

Model design

In total 5 random effects models (from 5 synthetic data types) were tested:

- only intercept, M = 10 (centered, non-centered),
- multivariate regression, M = 10 (centered, non-centered),
- multivariate regression, M per individual (centered, non-centered),
- no known process (centered, non-centered),
- multivariate regression with interaction, M per individual (centered, non-centered),

```
transformed parameters{
    vector[I] SI;          // SI index (per child)
    vector[I] Ht;          // true entropy (per chi

    // linear predictor
    for(i in 1:I){
      SI[i] = re_i[i] + a + aHS[HS[i]] +
        bA*Am[i] + bP*sPTA[i];
      // no multicollinearity between E and HS
    }

    // average entropy (SI -> Ht: negative)
    Ht = inv_logit(-SI);  // average entropy (SI ->
}
model{

    // hyperpriors
    m_i ~ normal( 0 , 0.2 );
    s_i ~ exponential( 1 );

    // priors
    a ~ normal( 0 , 0.2 );
    re_i ~ normal( m_i , s_i );
    //aE ~ normal( 0 , 0.5 );
    aHS ~ normal( 0 , 0.5 );
    bP ~ normal( 0 , 0.3 );
    bA ~ normal( 0 , 0.3 );
    m_M ~ lognormal( 1.5 , 0.5 );

    // likelihood
    for(n in 1:N){
      H[n] ~ beta_proportion( Ht[cid[n]] , m_M );
    }
}
```

Model design

Procedure:

- notice we used the hypothesis and (some) probabilistic assumptions defined in section Estimand and process model
- is like running section Synthetic data generation backwards

```
transformed parameters{
    vector[I] SI;          // SI index (per child)
    vector[I] Ht;          // true entropy (per child)

    // linear predictor
    for(i in 1:I){
      SI[i] = re_i[i] + a + aHS[HS[i]] +
        bA*Am[i] + bP*sPTA[i];
      // no multicollinearity between E and HS
    }

    // average entropy (SI -> Ht: negative)
    Ht = inv_logit(-SI);  // average entropy (SI -> H
}
model{
    // hyperpriors
    m_i ~ normal( 0 , 0.2 );
    s_i ~ exponential( 1 );
    m_M ~ normal( 0 , 0.5 );
    s_M ~ exponential( 1 );

    // priors
    a ~ normal( 0 , 0.2 );
    re_i ~ normal( m_i , s_i );
    M ~ lognormal( m_M , s_M );
    //aE ~ normal( 0 , 0.5 );
    aHS ~ normal( 0 , 0.5 );
    bP ~ normal( 0 , 0.3 );
    bA ~ normal( 0 , 0.3 );

    // likelihood
    for(n in 1:N){
      H[n] ~ beta_proportion( Ht[cid[n]] , M[cid[n]]
    }
}
```

Prior predictive simulation

Priors and hyper-priors

- In the probabilistic (causal) model there were no priors for our parameters,
- To decide our priors we follow McElreath [7]: "priors are part of the assumptions, and should be inspected as such",
- We will evaluate the implications of our priors on the outcome scale. We have three outcomes scales: SI_i, H_i^T, and H_{ik}^O

Priors

$$
\begin{aligned}
a_i &\sim \text{Normal}(\mu_a, \sigma_a) \\
M_i &\sim \text{LogNormal}(\mu_M, \sigma_M) \\
\alpha &\sim \text{Normal}(0, 0.2) \\
\alpha_{HS[i]} &\sim \text{Normal}(0, 0.5) \\
\alpha_{E[i]} &\sim \text{Normal}(0, 0.5) \\
\beta_{A,HS[i]} &\sim \text{Normal}(0, 0.3) \\
\beta_P &\sim \text{Normal}(0, 0.3)
\end{aligned}
$$

Hyper-priors

$$
\begin{aligned}
\mu_a &\sim \text{Normal}(0, 0.2) \\
\sigma_a &\sim \text{Exp}(1) \\
\mu_M &\sim \text{Normal}(0, 0.5) \\
\sigma_M &\sim \text{Exp}(1)
\end{aligned}
$$

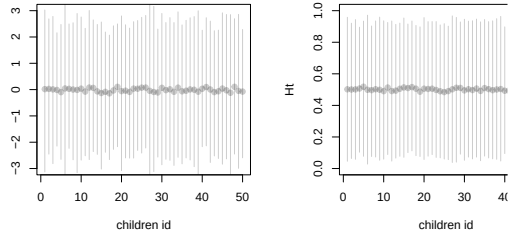
Prior predictive simulation

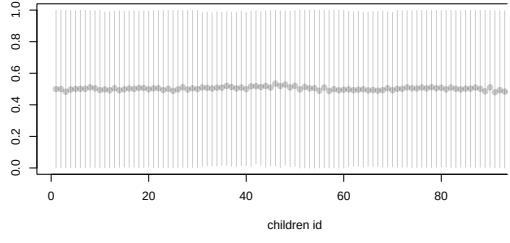
What our priors imply?

NO undesired assumption has crept in:

- the SI_i scale,
- the H_i^T scale,
- the H_{ik}^O scale

i.e. the full space of the scales can be reached by the parameters





Parameter recovery

Posterior predictive

Power

2. Research hypothesis procedure

Apply statistical model to data

What is going on?

3. References

3. References

[1] Boonen, N., Kloots, H., Nurzia, P. and Gillis, S. [2021]. Spontaneous speech intelligibility: early cochlear implanted children versus their normally hearing peers at seven years of age, Journal of Child Language pp. 1–26.
doi: https://doi.org/10.1017/S0305000921000714.

[2] De Raeve, L. [2016]. Cochlear implants in belgium: Prevalence in paediatric and adult cochlear implantation, European Annals of Otorhinolaryngology, Head and Neck Diseases 133: S57–S60.
doi: https://doi.org/10.1016/j.anorl.2016.04.018.
url: https://www.sciencedirect.com/science/article/pii/S1879729616300813.

[3] Faes, J., De Maeyer, S. and Gillis, S. [2021]. Speech intelligibility of children with an auditory brainstem implant: a triple-case study, pp. 1–50. (submitted).

[4] Fogarty, L., Madeleine, A., Holding, T., Powell, A. and Kandler, A. [2022]. Ten simple rules for principled simulation modelling, PLOS Computational Biology 18(3): 1–8.
doi: https://doi.org/10.1371/journal.pcbi.1009917.

[5] Hernán, M. [2020]. Causal diagrams: Draw your assumptions before your conclusions.
url: https://www.edx.org/course/causal-diagrams-draw-your-assumptions-before-your.

[6] Kruschke, J. [2014]. Doing Bayesian Data Analysis, A Tutorial with R, JAGS, and Stan, Elsevier.

[7] McElreath, R. [2020]. Statistical Rethinking: A Bayesian Course with Examples in R and STAN, Chapman and Hall/CRC.

[8] McElreath, R. [2022]. Statistical rethinking, 2022 course.
url: https://github.com/rmcelreath/stat_rethinking_2022.

[9] Pearl, J. [2019]. The seven tools of causal inference, with reflections on machine learning, Communications of the ACM 62(3): 54–60.
doi: https://doi.org/10.1177/0962280215586010.