Données en sciences cognitives
Chronométrie mentale
Cours 2

Gabriel Weindel
gabriel.weindel@univ-amu.fr

21 janvier 2020

Plan

1. Simulating 2AFC reaction times
   - Reminder
   - Simulating a single participant
   - Simulating multiple participants
   - Generating the dataset
2. Linear models
   - Visualisation of RT data
   - Reminder
   - Application to the data
3. Linear Mixed Models
   - Formalisation
   - Principle of random effects

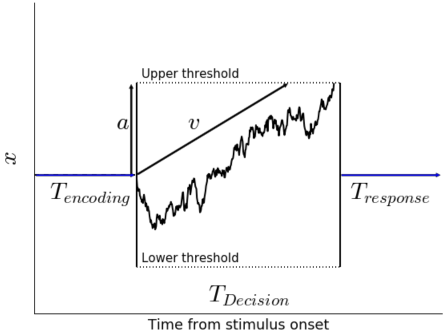
Plan

1. Simulating 2AFC reaction times
   - Reminder
   - Simulating a single participant
   - Simulating multiple participants
   - Generating the dataset
2. Linear models
   - Visualisation of RT data
   - Reminder
   - Application to the data
3. Linear Mixed Models
   - Formalisation
   - Principle of random effects

Simulating 2AFC reaction times:

Reminder

Random walk/diffusion models (Stone, 1960; Ratcliff, 1978), applicable to RT data from choice RT tasks



Wagenmakers, van der Maas et Grasman (2007) :

$$TRM = TDM + T_{er} \quad (1)$$

$$TDM = \left(\frac{a}{2v}\right) \frac{1 - e^{-va/s^2}}{1 + e^{-va/s^2}} \quad (2)$$

$$P_u = \frac{1}{1 + e^{-av/s^2}} \quad (3)$$

Simulating 2AFC reaction times:

Reminder

In a process model, parameters are used to infer on the (assumed) cognitive processes, e.g. in the diffusion model

- accumulation speed (v) $\rightarrow$ processing speed of the stimulus
- decision boundary (a) $\rightarrow$ response precaution
- Non-decision time (T_{er}) $\rightarrow$ residual time (e.g. stimulus encoding and response execution)

We can therefore test IVs on the parameters of interest (e.g. does age reduce processing speed of the stimulus)

Simulating 2AFC reaction times:

Simulating a single participant

Notebook

Load notebook 1

Simulating 2AFC reaction times:

Simulating a single participant

The (simple) Drift Diffusion model (DDM) has a known density function (Danielle J. Navarro & Ian G. Fuss, 2009).

Hence given parameter values ($v = .1$, $a = 1$, $t_{er} = 0.3$) we can draw the associated density function

Notebook

See cell[1:5] in notebook 1

When we are performing an experiment/a simulation, we are drawing from this distribution.
We can therefore compare the theoretical density to the samples to determine for example sample size, precision, sensitivity of descriptive statistics,. . .

Notebook

See cell[6:10] in notebook 1 for an example of two sampling and the resulting mean RT

We can draw a histogram of the simulated participant and look at the effect of changing the parameters values

Notebook

See cell[11:14] in notebook 1 for examples of RT histograms

Simulating 2AFC reaction times:

Simulating a single participant

Exercise

- Change the number of bins on the histogram, try lower vs higher numbers of bins. What do you see ?

Simulating 2AFC reaction times:

Simulating a single participant

Exercise

- Change the number of bins on the histogram, try lower vs higher numbers of bins. What do you see ?
- Create a participant that execute responses slower than the one simulated here and plot the histograms to compare

Simulating 2AFC reaction times:

Simulating multiple participants

Notebook

Load notebook 2

Simulating 2AFC reaction times:

Simulating multiple participants

We want to create a group of 50 participants which presents inter-individual variability. Hence, each individual should receive a unique value for the three parameters of the simple DDM.

But how do we define this inter-individual variability ?

Simulating 2AFC reaction times:

Simulating multiple participants

We want to create a group of 50 participants which presents inter-individual variability. Hence, each individual should receive a unique value for the three parameters of the simple DDM.

But how do we define this inter-individual variability ? By defining

distributions on the parameters !

Notebook

See cell[2] to see the distributions of parameters we will sample from
See cell[3:5] for the sampling

Simulating 2AFC reaction times:

Simulating multiple participants

Adding an experimental condition like we did to simulate a slower participant, except we still have to generate some inter-individual variability.

Notebook

See cell[6:8] for a new condition of Speed Accuracy tradeoff (SAT)

Simulating 2AFC reaction times:

Generating the dataset

Now we are going to create a *within* factorial design with two IVs for the 50 participants :

- Easiness with 5 levels (e.g. Word frequency, contrast level, ...)
- Quality of the stimulus, intact vs. degraded

Notebook

See cell[9:11] for a new condition on the stimulus

Simulating 2AFC reaction times:

Generating the dataset

Exercise

- Replace the second IV by a SAT condition with two levels as seen previously by changing the for loop (do not save the data)
- Summarize the distribution of the parameter in the two conditions (mean and sd in modality speed vs standard)
- Plot the histogram of simulated RT between both conditions at the group level

Plan

Linear models:

Visualisation of RT data

Notebook

Load notebook 3
See cell[1:8] for a way to plot a 2 $\times$ 2 factorial design on mean RT

Linear models:

Visualisation of RT data

Notebook

Load notebook 3
See cell[1:8] for a way to plot a 2×2 factorial design on mean RT

Don't forget that other descriptive statistics are available, for simplicity we use mean RT but we could use other statistics (e.g. median, quantiles)

Linear models:

Reminder

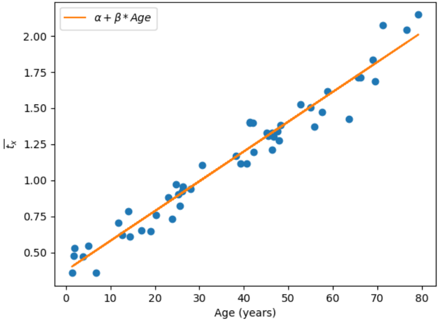
Linear Models are extremely useful statistical tool, they describe linear relationship between the IV (e.g. Age, continous) and a measure (e.g. $\overline{T}$) :

$$\overline{t_x} \sim \mathcal{N}(\mu, \sigma)$$

$$\mu = \alpha + \beta * IV$$

Linear models:
Reminder

Linear Models are extremely useful statistical tool, they describe the linear relationship between the IV (e.g. Age, continous) and a measure (e.g. $\overline{T}$) :



Linear models:

Application to the data

Now we are going to see how we can apply a simple linear model on the mean RT we just extracted

FYI : What we are going to do is for illustrative purpose, we cannot perform a linear model on data that is nested (i.e. within design) because a linear regression assumes that data points are independent (which they are not as the same participant is tested in 5 stimulus difficulty levels and in the two speed conditions)

Linear models:

Application to the data

First we are computing the mean RT ($\bar{t}_x j$) for each experimental cell xj (participant x and each modality combination j).

Then we use the *lm*() R command to estimate the parameters of the linear model (α, β_1, β_2, β_{12}).

Notebook

See cell[9:13] for a LM using the easiness factor
See cell[14:19] for a LM using the encoding factor
See cell[20:22] for a multiple LM using the encoding and easiness factor
See cell[23:25] for a multiple LM using the encoding and easiness factor with interaction

Linear models:

Application to the data

All statistical models come with assumption (e.g. independence, normality of the residuals, homeoscedasticity) we can check some of these assumption after the fit.

Notebook

See cell[26:27] for a check of the assumption on the last LM

Linear models:

Application to the data

Exercise

- Reproduce the linear model with interaction by prealably transforming the RT to log(RT), knowing that the R command for log-transform is log(x) where x is the variable of interest
- Reproduce the linear model with interaction but by changing the coding of easiness and taking the first modality as the reference condition

Plan

Linear Mixed Models describe linear relationship between the IV (e.g. Age, continous) and a measure (e.g. $\overline{t}_x j$) by accounting for the hierarchy of the data (i.e. multiple observation for each participant x):

$$\overline{t}_{xj} \sim \mathcal{N}(\mu, \sigma)$$

$$\mu = \alpha_x + \beta_{ix} * IV_i$$

where random effects are defined as follow :

$$\alpha \sim \mathcal{N}(\mu_\alpha, \sigma_\alpha)$$
$$\beta_i \sim \mathcal{N}(\mu_{\beta_i}, \sigma_{\beta_i})$$

Linear Mixed Models:

Principle of random effects

Notebook

Load notebook 4

Linear Mixed Models:

Principle of random effects

Exercise

- Reproduce the linear model from the previous section with only the easiness predictor and the same subsampled data $data[(data.participant < 6)\&(data.encoding == "standard"),]$
- Compare the summary between the simple linear model and the last mixed model (M3)