Explanatory models

Joachim Vandekerckhove

Types of hierarchical models

The basic form of a hierarchy in a parametric model is:

$$\mathcal{M}_s : \begin{cases} \forall p \in (1, \dots, P) : \mu_p \sim N(M, T) \\ \forall j \in (1, \dots, J), p \in (1, \dots, P) : x_{p,j} \sim N(\mu_p, \tau) \\ M \sim N(0, 0.1) \\ \tau \sim \Gamma(4, 0.01) \\ T \sim \Gamma(4, 0.01) \end{cases}$$

The hierarchical part is $\mu_p \sim N(M, T)$, but this is only one possible hierarchy.

Difference structure

Consider one alternative:

$$\mu_{p|g} \sim N(M_g, T)$$

Here the parameter M is determined by group identity g, and person p, who is a member of g, has some expected parameter value based on that.

If there are two groups, we may be interested in between-group differences $D_{21} = M_2 - M_1$.

Regression structure

The difference structure $D_{21} = M_2 - M_1$ is a very basic form of linear regression. The difference structure is equivalent to this:

$$\mu_p \sim N(M_1 + z_p D_{21}, T)$$

where $z_p = 0$ if p is in group 1 and $z_p = 1$ if p is in group 2.

This way, z acts as a linear predictor of the person-specific hierarchical mean.

Person-specific hierarchical mean

Hierarchical means can belong to an individual.

We then interpret it as the mean of the population of individuals who have predictors similar to this one.

But it is possible (and common) to have a hierarchical mean that is not shared with any other participant in our sample.

Consider that the predictor z could be not group membership, but some continuous predictor like age, personality, etc.

Person-specific hierarchical mean

In the case where z is a property of the person, and

$$\mu_p \sim N(\beta_0 + \beta_1 z_p, T)$$

... we have a standard linear regression structure with intercept β_0 and regression weight β_1.

Person side and item side predictors

Suppose that there are other sources of variability in the data – the difficulty of items i in a task might vary as well. Then,

$$\mu_{p,i} \sim N(\beta_0 + \beta_1 z_p + \beta_2 c_i, T)$$

Here z_p would be called a "person-side" predictor and c_i an "item-side" predictor.

You could also have an "interaction side."

Plug then play

Critically, nothing in this framework depends on data being (conditionally) normally distributed.

So far we have "decomposed" the person-specific hierarchical mean, which is a parameter of the normal distribution:

$$\mu_{p,i} \sim N(\beta_0 + \beta_1 z_p + \beta_2 c_i, T)$$

Plug then play

If your data are Bernoulli distributed (say, participants getting items right or wrong), maybe you want to decompose the success rate parameter $\pi_{p,i}$ into a person ability parameter θ_p and an item difficulty parameter β_i:

$$\begin{cases} x_{p,i} \sim \text{Bernoulli}(\pi_{p,i}) \\ \pi_{p,i} = \text{ilogit}\left(\eta_{p,i}\right) \\ \eta_{p,i} = \theta_p - \beta_i \\ \beta_i \sim N(0, T_\beta) \\ \theta_p \sim N(M, T_\theta) \end{cases}$$

Linking functions

Note that we're interested in decomposing the success rate, $\pi_{p,i}$, but instead work on a different parameter, $\eta_{p,i}$.

We do that because $\pi_{p,i}$ (a probability) cannot easily be linearly decomposed, since it cannot be less than 0 or greater than 1

The transformed parameter $\eta_{p,i}$ lives in the domain of the real numbers and thus can be linearly decomposed.

In order to link $\eta_{p,i}$ to a probability, we need to map the parameter to the $(0, 1)$ domain, where probabilities live.

For this, we will use a linking function – a function that takes as input $\eta_{p,i} \in \mathbb{R}$ and gives as output a value $(0, 1)$.

$$\mathbb{R} \xrightarrow{\text{link}} (0, 1)$$

Linking functions

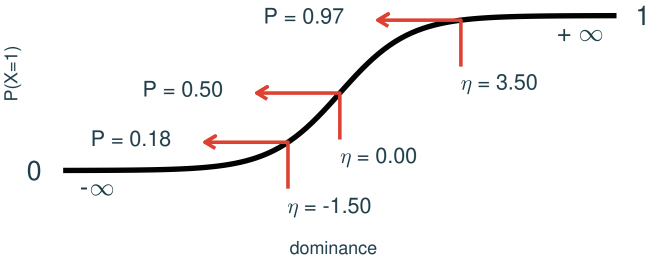
Which linking function you use depends on context (what is the valid range of the natural parameter?) and on culture.

Psychometricians will usually use the logistic function (also known as the inverse logit):

$$\pi_{p,i} = \frac{1}{1 + \exp(-\eta_{p,i})}$$

Which linking function you use depends on context (what is the valid range of the natural parameter?) and on culture. Psychometricians will usually use the logistic function (also known as the inverse logit):

$$\pi_{p,i} = \frac{1}{1 + \exp(-\eta_{p,i})}$$



The Rasch model

A much more convenient way of writing that is:

$$\text{ilogit}\left(\eta_{ip}\right) = \text{ilogit}\left(\theta_p - \beta_i\right)$$

$$\Leftrightarrow \quad \text{logit}\left(\pi_{p,i}\right) = \theta_p - \beta_i$$

This exact formulation is one of the most common item response models (there are a few). This is called the Rasch model, after Danish psychometrician Georg Rasch (Rasch 1961).

(We're not getting into that.)

Plug then play

The same logic can apply to more “exotic” distributions as well.

So far we have decomposed the person-specific hierarchical mean (of a normal distribution) and the person-specific hierarchical probability of success (of a Bernoulli distribution).

You can decompose the parameters of an arbitrary distribution $\mathcal{D}()$:

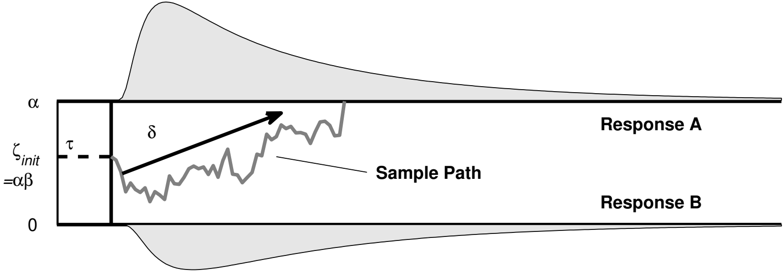
$$X_* \sim \mathcal{D}\left(\theta_*, Z_*, \beta\right)$$

... for some data X_*, predictors Z_*, and hierarchical parameter vector β.

Let’s explore a more exotic distribution.

Data analysis with diffusion models

	parameter	*interpretation*
δ	drift rate	dominance (η, d')
α	boundary separation	caution
τ	nondecision time	time for encoding and responding
β	initial bias	a priori response bias



Diffusion model parameter estimation

person p	*condition* c	*RT*	*accuracy*
1	3	0.71	correct
1	5	0.49	correct
⋮	⋮	⋮	⋮
1	3	0.43	error
2	4	0.67	error
⋮	⋮	⋮	⋮
9	2	0.61	correct
9	2	0.39	error

α_p	δ_{pc}	τ_p
1.61	0.45	0.24
1.61	1.17	0.24
⋮	⋮	⋮
1.61	0.53	0.24
2.14	0.08	0.31
⋮	⋮	⋮
1.41	0.79	0.27
1.41	0.79	0.27

Often, we will perform constrained parameter estimation so (e.g.) each participant has one α_p and τ_p, independent of condition

An explanatory diffusion model

In the explanatory item response model, we decomposed the parameter η into a mathematical expression that allows for structured differences between items, persons, etc.

By analogy, we could make an explanatory diffusion model, where the drift rate δ is given the same treatment

$$p(t_{p,c}, x_{p,c}) = W(\delta_{p,c}, \alpha_p, \tau_p, \beta_p)$$

$$= W(\theta_p + \lambda_c, \alpha_p, \tau_p, \beta_p)$$

$$= W(\theta_p + \gamma Z_c, \alpha_p, \tau_p, \beta_p)$$

An explanatory diffusion model account of shape perception

We used this model to revisit data from (Vandekerckhove, Panis, and Wagemans 2007) (a shape perception study)

An explanatory diffusion model account of shape perception

We used this model to revisit data from (Vandekerckhove, Panis, and Wagemans 2007) (a shape perception study)

We decomposed the drift rate parameter to account for three independent variables in the experiment:

An explanatory diffusion model account of shape perception

We used this model to revisit data from (Vandekerckhove, Panis, and Wagemans 2007) (a shape perception study)

We decomposed the drift rate parameter to account for three independent variables in the experiment:

- $Q_c = 1$ for trials with a quantitative change, $Q_c = 0$ otherwise

An explanatory diffusion model account of shape perception

We used this model to revisit data from (Vandekerckhove, Panis, and Wagemans 2007) (a shape perception study)

We decomposed the drift rate parameter to account for three independent variables in the experiment:

- $Q_c = 1$ for trials with a quantitative change, $Q_c = 0$ otherwise
- $V_c = 1$ for trials with a concave change, $V_c = 0$ otherwise

An explanatory diffusion model account of shape perception

We used this model to revisit data from (Vandekerckhove, Panis, and Wagemans 2007) (a shape perception study)

We decomposed the drift rate parameter to account for three independent variables in the experiment:

- $Q_c = 1$ for trials with a quantitative change, $Q_c = 0$ otherwise
- $V_c = 1$ for trials with a concave change, $V_c = 0$ otherwise
- $N_c = 1$ for trials with no change, $N_c = 0$ otherwise

An explanatory diffusion model account of shape perception

We used this model to revisit data from (Vandekerckhove, Panis, and Wagemans 2007) (a shape perception study)

We decomposed the drift rate parameter to account for three independent variables in the experiment:

- $Q_c = 1$ for trials with a quantitative change, $Q_c = 0$ otherwise
- $V_c = 1$ for trials with a concave change, $V_c = 0$ otherwise
- $N_c = 1$ for trials with no change, $N_c = 0$ otherwise

An explanatory diffusion model account of shape perception

We used this model to revisit data from (Vandekerckhove, Panis, and Wagemans 2007) (a shape perception study)

We decomposed the drift rate parameter to account for three independent variables in the experiment:

- $Q_c = 1$ for trials with a quantitative change, $Q_c = 0$ otherwise
- $V_c = 1$ for trials with a concave change, $V_c = 0$ otherwise
- $N_c = 1$ for trials with no change, $N_c = 0$ otherwise

We also set $\beta = 1/2$

An explanatory diffusion model account of shape perception

We used this model to revisit data from (Vandekerckhove, Panis, and Wagemans 2007) (a shape perception study)

We decomposed the drift rate parameter to account for three independent variables in the experiment:

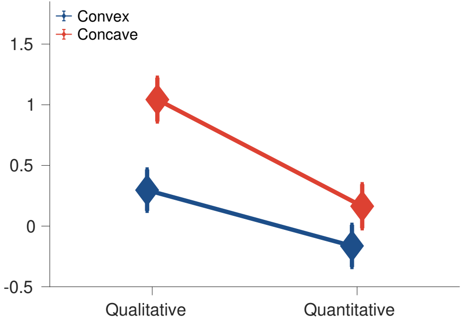
- $Q_c = 1$ for trials with a quantitative change, $Q_c = 0$ otherwise
- $V_c = 1$ for trials with a concave change, $V_c = 0$ otherwise
- $N_c = 1$ for trials with no change, $N_c = 0$ otherwise

We also set $\beta = {}^1\!/_2$

$$p(t_{p,c}, x_{p,c}) = W(\theta_p + \lambda_c, \alpha_c, \tau_c, {}^1\!/_2)$$

$$= W(\theta_p + \mu + \gamma_1 Q_c + \gamma_2 V_c + \gamma_3 Q_c V_c + \gamma_4 N_c, \alpha_c, \tau_c, {}^1\!/_2)$$

Process models of speeded choice response time



	Posterior	
	Mean	SD
μ	0.304	0.174
γ_1	-0.464	0.247
γ_2	0.729	0.243
γ_3	-0.408	0.341
γ_4	0.936	0.246

$$p(t_{p,c}, x_{p,c}) = W(\theta_p + \mu + \gamma_1 Q_c + \gamma_2 V_c + \gamma_3 Q_c V_c + \gamma_4 N_c, \alpha_p, \tau_p, {}^1\!/_2)$$

(Note: Find the rest of the analysis on `osf.io/hf96a`.)

Random effects, manifest, latent

We have now seen two kinds of hierarchy:

- Random effects: There are individual differences but everyone is from the same population

Random effects, manifest, latent

We have now seen two kinds of hierarchy:

- Random effects: There are individual differences but everyone is from the same population
- Regression style: There are individual differences and they are party explained by observed predictors

Random effects, manifest, latent

We have now seen two kinds of hierarchy:

- Random effects: There are individual differences but everyone is from the same population
- Regression style: There are individual differences and they are party explained by observed predictors

Random effects, manifest, latent

We have now seen two kinds of hierarchy:

- Random effects: There are individual differences but everyone is from the same population
- Regression style: There are individual differences and they are party explained by observed predictors

Observed predictors are often called manifest.

Random effects, manifest, latent

We have now seen two kinds of hierarchy:

- Random effects: There are individual differences but everyone is from the same population
- Regression style: There are individual differences and they are party explained by observed predictors

Observed predictors are often called manifest.

Predictors may be unobserved, in which case we call them latent.

A cognitive latent variable model

Suppose that participants perform multiple tasks, and we expect their drift rate to be affected by their (unobserved) ability to inhibit errors, E_p and to direct attention A_p.

$$\delta_{p,i} = \lambda_0 + \lambda_1 E_p + \lambda_2 A_p - \beta_i$$

A cognitive latent variable model

Suppose that participants perform multiple tasks, and we expect their drift rate to be affected by their (unobserved) ability to inhibit errors, E_p and to direct attention A_p.

$$\delta_{p,i} = \lambda_0 + \lambda_1 E_p + \lambda_2 A_p - \beta_i$$

This looks very similar (and it is) but remember that E and A are unobserved and so have to be estimated. This complicates things a lot, but it is possible (Vandekerckhove 2014).

Hierarchical model building blocks

- Marginal distribution: Distribution of the data given all parameters (ideally based on relevant theory)

Hierarchical model building blocks

- Marginal distribution: Distribution of the data given all parameters (ideally based on relevant theory)
- Link functions: Optionally, to map range-restricted parameters to the real line

Hierarchical model building blocks

- Marginal distribution: Distribution of the data given all parameters (ideally based on relevant theory)
- Link functions: Optionally, to map range-restricted parameters to the real line
- Hierarchy: Distributions of individual-level parameters given higher-order parameters (usually just normals)

Hierarchical model building blocks

- Marginal distribution: Distribution of the data given all parameters (ideally based on relevant theory)
- Link functions: Optionally, to map range-restricted parameters to the real line
- Hierarchy: Distributions of individual-level parameters given higher-order parameters (usually just normals)
- Structure: Statistical relationships between higher-order model parameters

Hierarchical model building blocks

- Marginal distribution: Distribution of the data given all parameters (ideally based on relevant theory)
- Link functions: Optionally, to map range-restricted parameters to the real line
- Hierarchy: Distributions of individual-level parameters given higher-order parameters (usually just normals)
- Structure: Statistical relationships between higher-order model parameters
  - Random effects: Parameters simply vary between persons

Hierarchical model building blocks

- Marginal distribution: Distribution of the data given all parameters (ideally based on relevant theory)
- Link functions: Optionally, to map range-restricted parameters to the real line
- Hierarchy: Distributions of individual-level parameters given higher-order parameters (usually just normals)
- Structure: Statistical relationships between higher-order model parameters
  - Random effects: Parameters simply vary between persons
  - Manifest predictors: Parameter variability is in part explained by observed predictors

Hierarchical model building blocks

- Marginal distribution: Distribution of the data given all parameters (ideally based on relevant theory)
- Link functions: Optionally, to map range-restricted parameters to the real line
- Hierarchy: Distributions of individual-level parameters given higher-order parameters (usually just normals)
- Structure: Statistical relationships between higher-order model parameters
  - Random effects: Parameters simply vary between persons
  - Manifest predictors: Parameter variability is in part explained by observed predictors
  - Latent predictors: Parameter variability is in part explained by unobserved predictors

Relevant recorded lectures

Item response theory

https://www.youtube.com/watch?v=dUjh2L1jy8I

Speeded response times

https://www.youtube.com/watch?v=AnuaLG5rYb8

Accumulator models of response time

https://www.youtube.com/watch?v=zgKAG0uoeDQ

Explanatory process models

https://www.youtube.com/watch?v=bkUnnm1E7sY

Cognitive latent variable models

https://www.youtube.com/watch?v=2XVd7bcXQ10

References

Rasch, Georg. 1961. "On General Laws and the Meaning of Measurement in Psychology." In Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, 4:321–33.

Vandekerckhove, Joachim. 2014. "A Cognitive Latent Variable Model for the Simultaneous Analysis of Behavioral and Personality Data." Journal of Mathematical Psychology 60: 58–71.

Vandekerckhove, Joachim, Sven Panis, and Johan Wagemans. 2007. "The Concavity Effect Is a Compound of Local and Global Effects." Perception & Psychophysics 69 (7): 1253–60. https://doi.org/10.3758/bf03193960.