Advanced Bayesian modeling

Joachim Vandekerckhove

Order of operations

$$X = \begin{bmatrix} 2 & 1 & 2 \\ 1 & 1 & 5 \end{bmatrix}$$

Order of operations

$$X = \begin{bmatrix} 2 & 1 & 2 \\ 1 & 1 & 5 \end{bmatrix}$$

Sum of sums:

$$\sum_{r=1}^{2} \left(\sum_{c=1}^{3} x_{rc} \right) = \sum_{c=1}^{3} \left(\sum_{r=1}^{2} x_{rc} \right)$$

Order of operations

$$X = \begin{bmatrix} 2 & 1 & 2 \\ 1 & 1 & 5 \end{bmatrix}$$

Sum of sums:

$$\sum_{r=1}^{2} \left(\sum_{c=1}^{3} x_{rc} \right) = \sum_{c=1}^{3} \left(\sum_{r=1}^{2} x_{rc} \right)$$

Sum of products vs. product of sums:

$$\sum_{r=1}^{2} \left(\prod_{c=1}^{3} x_{rc} \right) \neq \prod_{c=1}^{2} \left(\sum_{r=1}^{3} x_{rc} \right)$$

Order of operations

In general, order of operations matters:

$$f \circ g(x) \neq g \circ f(x)$$

Estimating parameters is an operation

In science, we often want to make statements about averages.

Estimating parameters is an operation

In science, we often want to make statements about averages.

Estimating model parameters from data is an operation:

$$\hat{\theta} = f(x)$$

Estimating parameters is an operation

In science, we often want to make statements about averages.

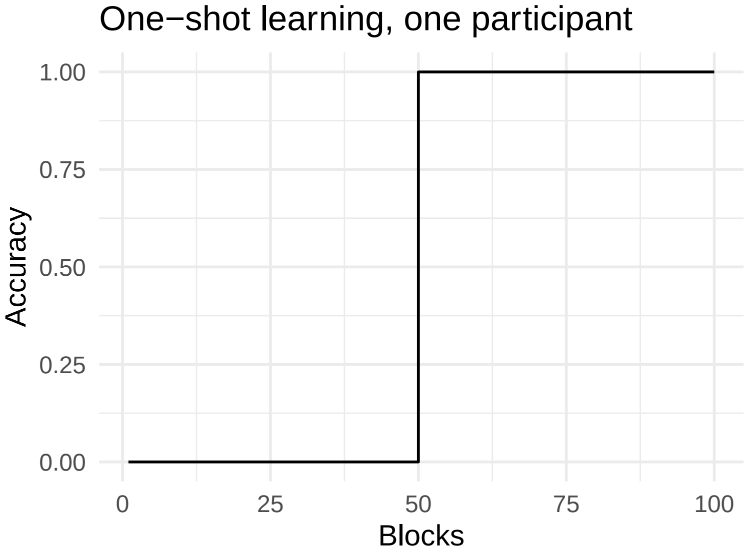
Estimating model parameters from data is an operation:

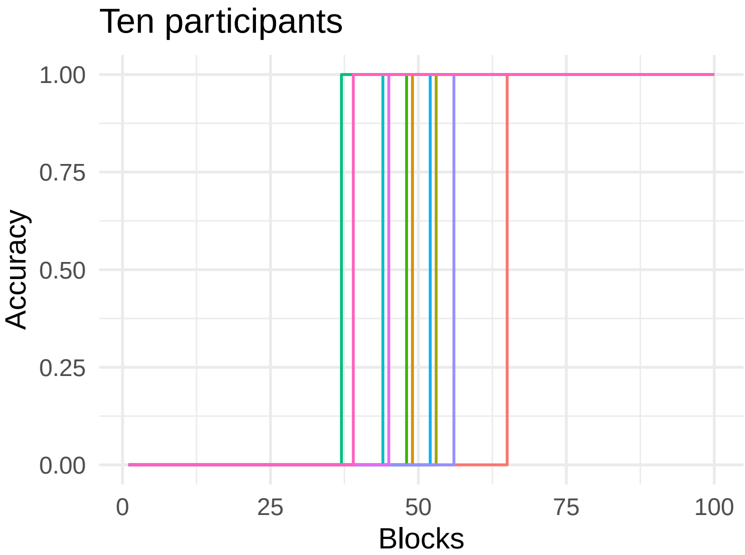
$$\hat{\theta} = f(x)$$

Do we want the average model parameters of the data or the model parameters of the average data?

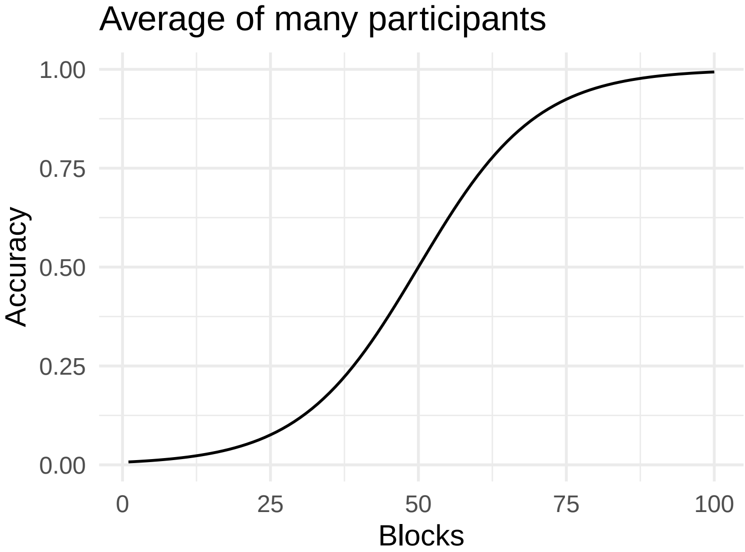
$$\overline{f(x)} \neq f\left(\overline{x}\right)$$

Insight learning



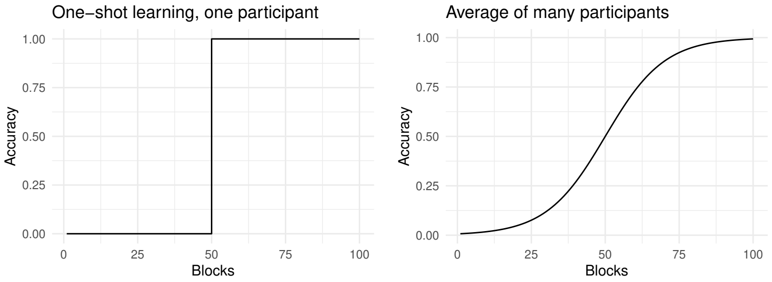

Ten participants
1.00
0.75
0.50
0.25
0.00
Accuracy
0
25
50
75
100
Blocks

Insight learning



Insight learning

The "average" learning curve looks nothing like the person-specific learning curve!



Insight learning

Instead, what we want here is to acknowledge that each person has their own trajectory, and then say something about the (average) properties of the trajectories.

Insight learning

Instead, what we want here is to acknowledge that each person has their own trajectory, and then say something about the (average) properties of the trajectories.

We want to make an abstraction of the data, which are something complicated that is generated by a process with parameters θ_p (for person p), and instead focus on parameters.

Basic hierarchy

$$\mathcal{M}_h : \begin{cases} x_p \sim \text{one-shot}\,(\theta_p) \\ \theta_p \sim N(\mu, \tau) \end{cases}$$

The hierarchical model contains one set of assumptions about the data given the model (the likelihood level), and another set of assumptions about structure among parameters.

Basic hierarchy

$$\mathcal{M}_h : \begin{cases} x_p \sim \text{one-shot}\,(\theta_p) \\ \theta_p \sim N(\mu, \tau) \end{cases}$$

The hierarchical model contains one set of assumptions about the data given the model (the likelihood level), and another set of assumptions about structure among parameters.

Here, the hierarchical parameters μ and τ tell us something about the population of participants, each with their own change point θ_p.

Population assumptions

An assumption made in hierarchical models is that it is possible to know things about participants merely because they are members of some sampled population.

Population assumptions

An assumption made in hierarchical models is that it is possible to know things about participants merely because they are members of some sampled population.

For example, just by knowing someone is a human, we can make reasonable estimates about their number of thumbs.

Population assumptions

An assumption made in hierarchical models is that it is possible to know things about participants merely because they are members of some sampled population.

For example, just by knowing someone is a human, we can make reasonable estimates about their number of thumbs.

If $n_t > 2$, a measurement error may have occurred.

Insight learning

Back to the one-shot learning, we could try to characterize the population of participants in terms of their change point distribution $\theta_p \sim N(\mu, \tau)$.

Insight learning

Back to the one-shot learning, we could try to characterize the population of participants in terms of their change point distribution $\theta_p \sim N(\mu, \tau)$.

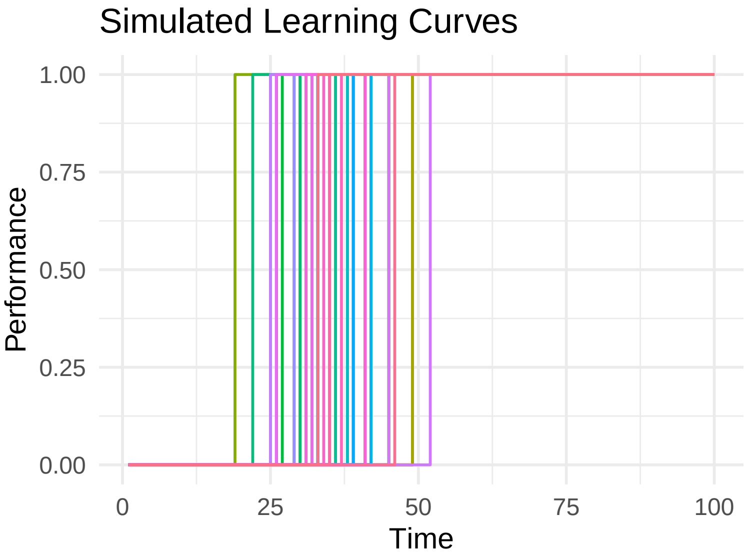
- Let's generate learning curves for a population of $n = 50$ participants, each with a random change point.

Insight learning

Back to the one-shot learning, we could try to characterize the population of participants in terms of their change point distribution $\theta_p \sim N(\mu, \tau)$.

- Let's generate learning curves for a population of $n = 50$ participants, each with a random change point.
- Let's use a step function to represent the learning curves.

Insight learning



Hierarchical model of insight learning

- We define a hierarchical model in JAGS to estimate each participant's change point.

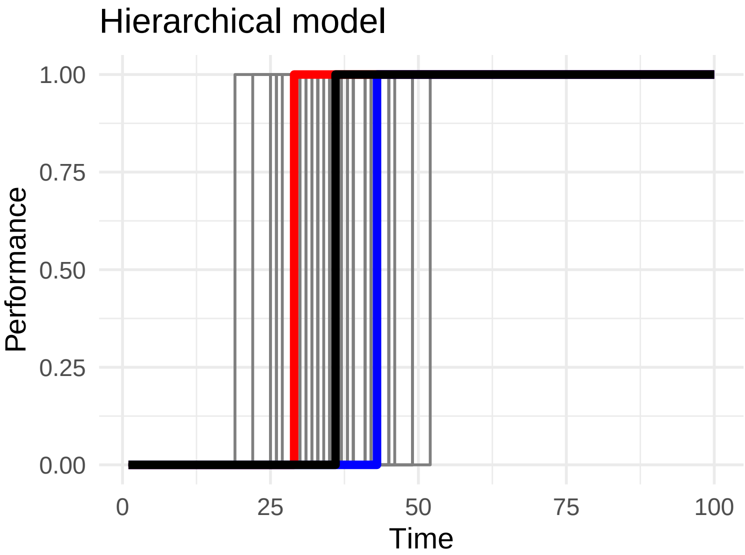
Hierarchical model of insight learning

- We define a hierarchical model in JAGS to estimate each participant's change point.
- The change point for each participant is modeled as a uniform distribution between 1 and 100.

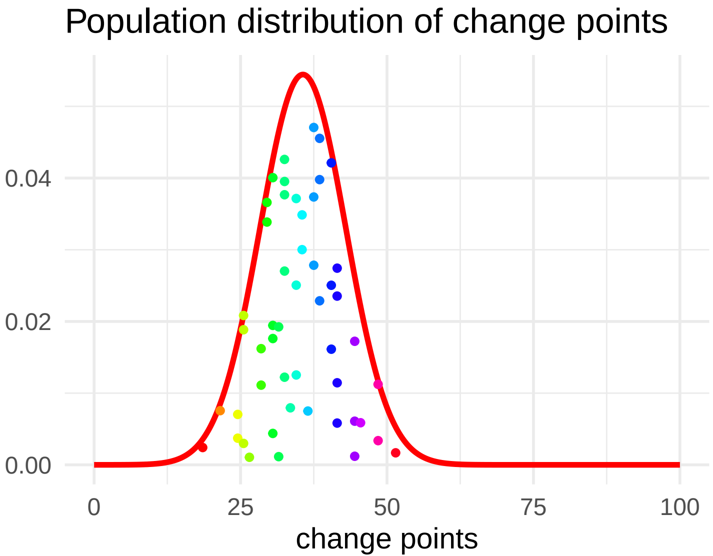
Hierarchical model of insight learning

```
##
## model {
##   for (j in 1:P) {
##     for (i in 1:N) {
##       y[i, j] ~ dbern(p[i, j])
##       p[i, j] <- ifelse(time[i] < theta[j], 0, 1)
##     }
##     theta[j] ~ dnorm(mu, tau)T(0,100)
##   }
##   mu  ~ dnorm(50, 0.05)T(0,100)
##   tau ~ dnorm(20, 0.10)T(0,)
##   sigma <- pow(tau, -0.5)
## }
```

Hierarchical model of insight learning



Population results



Parameter estimates

```
summary(samples_hrcl)[1]$
  statistics[c("mu","sigma"),
             c("Mean", "SD", "Time-series SE")] %>%
  print()
```

```
##              Mean        SD Time-series SE
## mu      35.642994 1.0194858    0.010932701
## sigma    7.328033 0.7454532    0.008005114
```

Parameter estimates

```r
summary(samples_hrcl)[1]$
  statistics[c("mu","sigma"),
             c("Mean", "SD", "Time-series SE")] %>%
  print()
```

```
##              Mean        SD Time-series SE
## mu      35.642994 1.0194858    0.010932701
## sigma    7.328033 0.7454532    0.008005114
```

Compare to:

- Simulated $\mu = 35$

Parameter estimates

```
summary(samples_hrcl)[1]$
  statistics[c("mu","sigma"),
             c("Mean", "SD", "Time-series SE")] %>%
  print()
```

```
##             Mean        SD Time-series SE
## mu     35.642994 1.0194858    0.010932701
## sigma   7.328033 0.7454532    0.008005114
```

Compare to:

- Simulated $\mu = 35$
- Simulated $\sigma = 8$