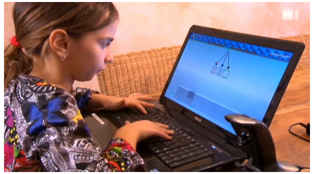
Knowledge Tracing

JDPLS – Theory Course

March 6, 2024

Tracing Student Knowledge

- Is the student learning?
  - Measure what the student *knows* at a specific time t
  - More specifically: knowledge of the student about relevant knowledge components (skills)



Task:	$50 - 23 = ?$	$75 - 12 = ?$	$38 - 14 = ?$
Answer:	27	61	24

Tracing Knowledge – why is it useful?

- Is the student learning?
  - Measure what the student *knows* at a specific time *t*
  - More specifically: knowledge of the student about relevant knowledge components (skills)

➡ Choose the next appropriate activity

➡ Know which activities support learning

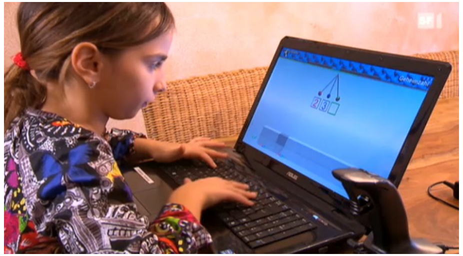
Tracing Knowledge – why is it useful?

- Is the student learning?
  - Measure what the student *knows* at a specific time *t*
  - More specifically: knowledge of the student about relevant knowledge components (skills)

➡ Choose the next appropriate activity

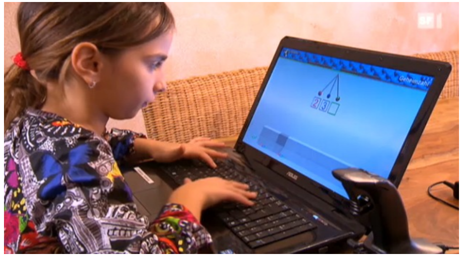
➡ Know which activities support learning

Is the student learning?



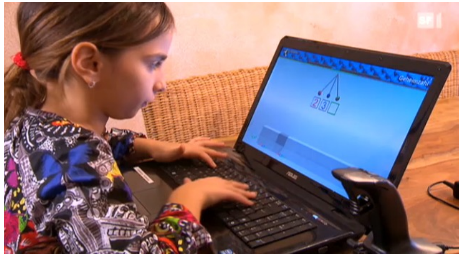
Task:	$50 - 23 = ?$	$75 - 12 = ?$	$38 - 14 = ?$
Answer:	27	61	24

What are we measuring?



Task:	$50 - 23 = ?$	$75 - 12 = ?$	$38 - 14 = ?$
Answer:	27	61	24
	1	0	1

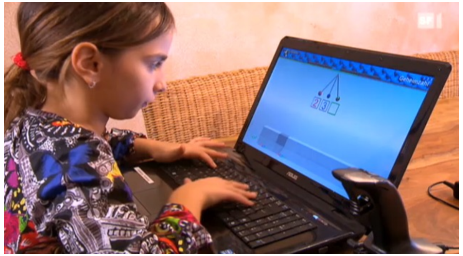
Binary observations of student answers



Subtraction 0-100

1	2		…		n
0	0	1	0	1	1

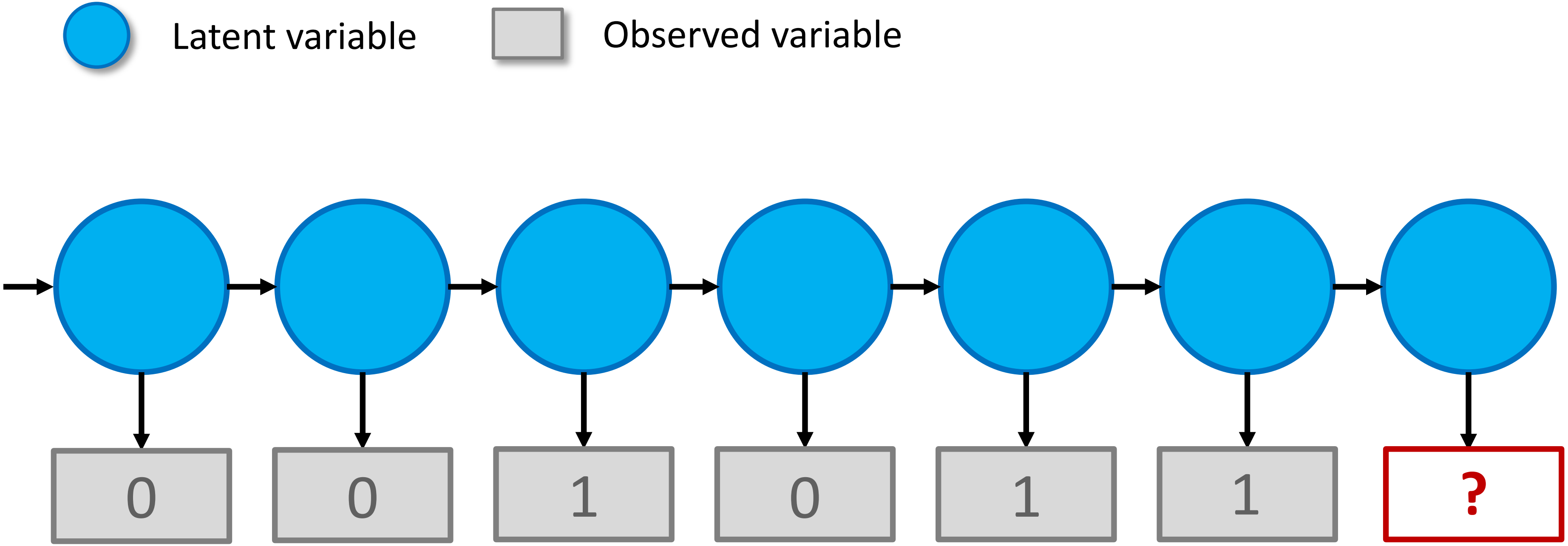
Predicting future performance



Subtraction 0-100

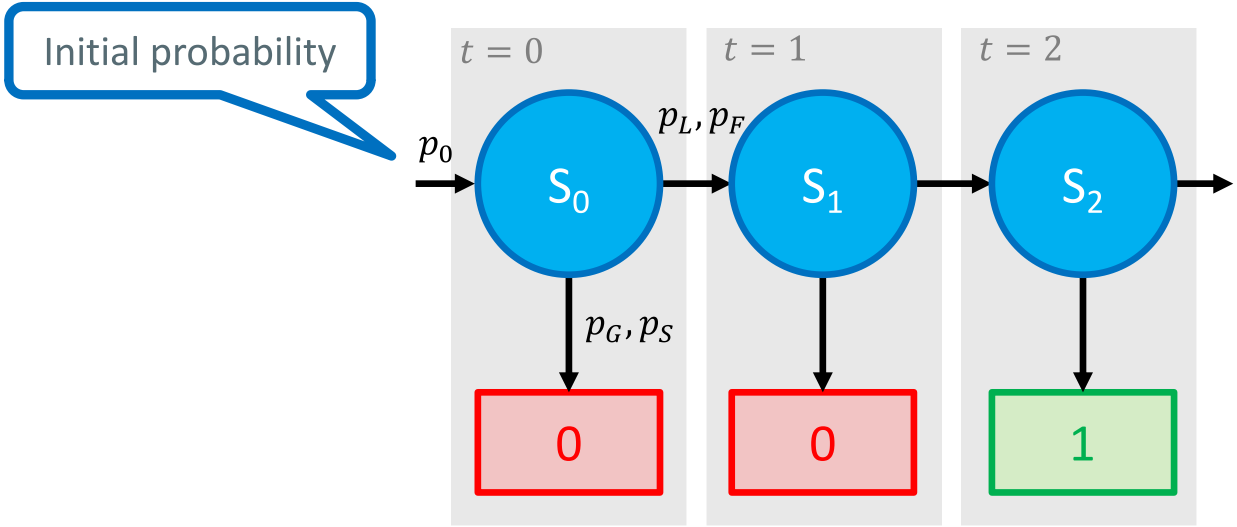
1	2		...		n	n+1
0	0	1	0	1	1	?

Bayesian Knowledge Tracing (BKT)

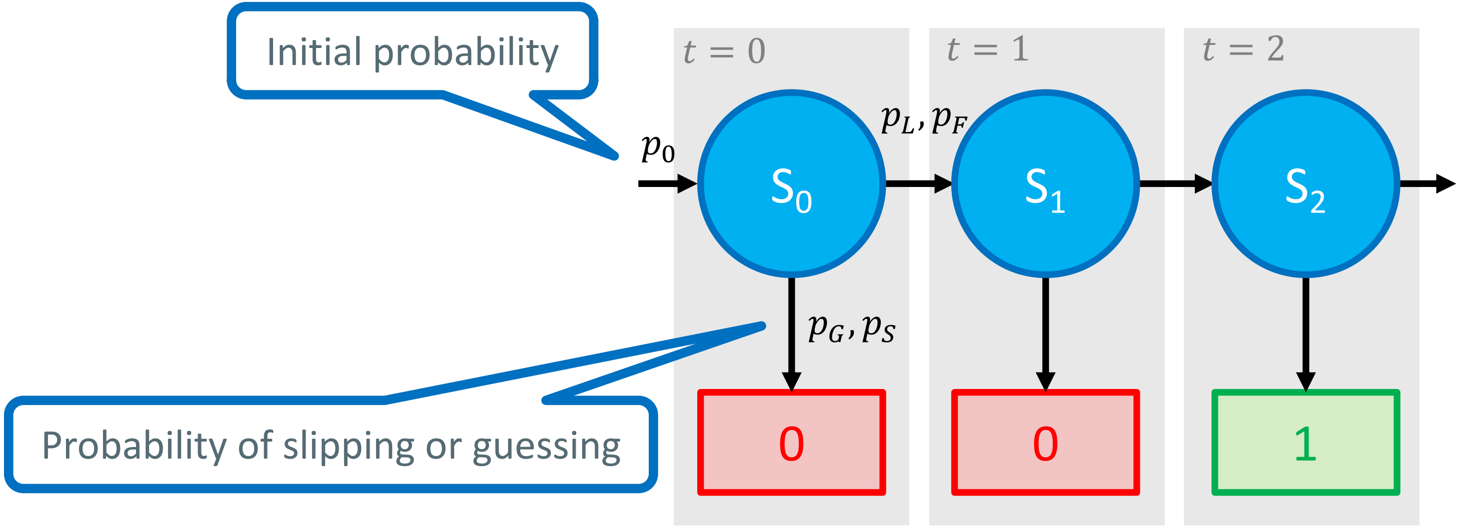


[Corbett & Anderson, UMUAI 1994]

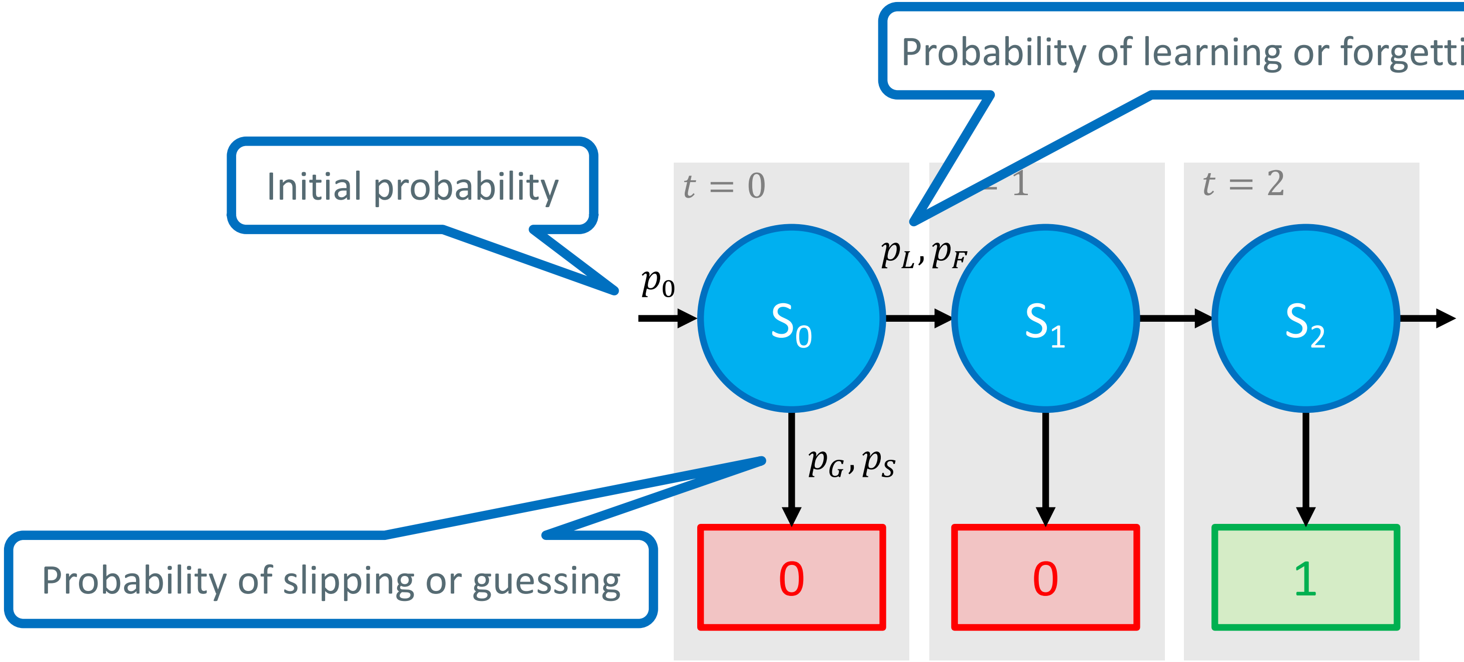
BKT parameters are interpretable



BKT parameters are interpretable



BKT parameters are interpretable



Two tasks need to be solved in practice

- Given a model with parameters $\theta = \{p_0, p_L, p_F, p_S, p_G\}$ and a sequence of observations $\boldsymbol{o} = [o_0, \dots, o_t]$ from a student s, predict o_{t+1}

o_0	o_1	o_2	$\bullet\bullet\bullet$		o_t	o_{t+1}
0	0	1	0	1	1	?

Inference

Making predictions using a BKT model

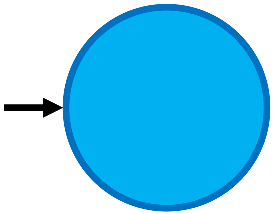
$p_0 = 0.5$

$p_S = 0.2$

$p_G = 0.3$

$p_L = 0.4$

$p_F = 0.0$



Making predictions using a BKT model

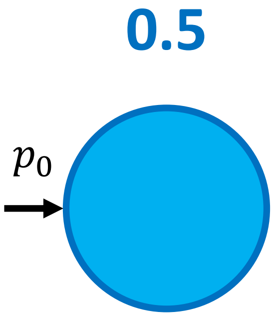
$p_0 = 0.5$

$p_S = 0.2$

$p_G = 0.3$

$p_L = 0.4$

$p_F = 0.0$



$p(s_0 = 1) = p_0$

Making predictions using a BKT model

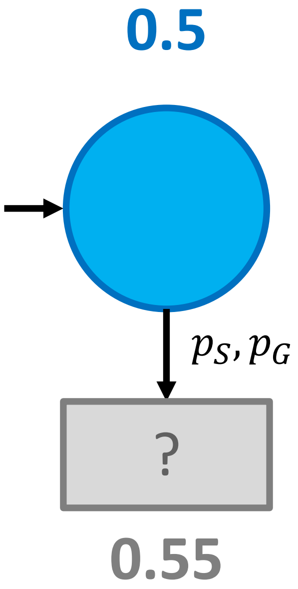
$p_0 = 0.5$

$p_S = 0.2$

$p_G = 0.3$

$p_L = 0.4$

$p_F = 0.0$



$$p(s_0 = 1) = p_0$$

$$p(o_0 = 1) = (1 - p_s) \cdot p_0 + p_G \cdot (1 - p_0)$$

Making predictions using a BKT model

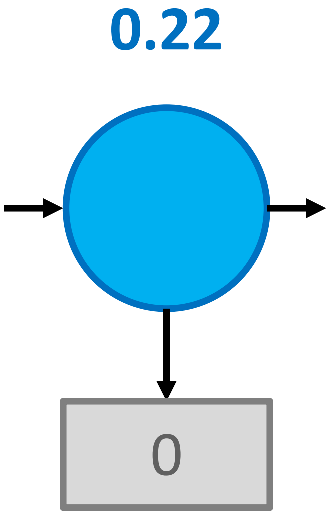
$p_0 = 0.5$

$p_S = 0.2$

$p_G = 0.3$

$p_L = 0.4$

$p_F = 0.0$



$$p_{s_0|0} = \frac{p_S \cdot p_0}{1 - p(o_0 = 1)} = \frac{0.2 \cdot 0.5}{0.45} = 0.22$$

Making predictions using a BKT model

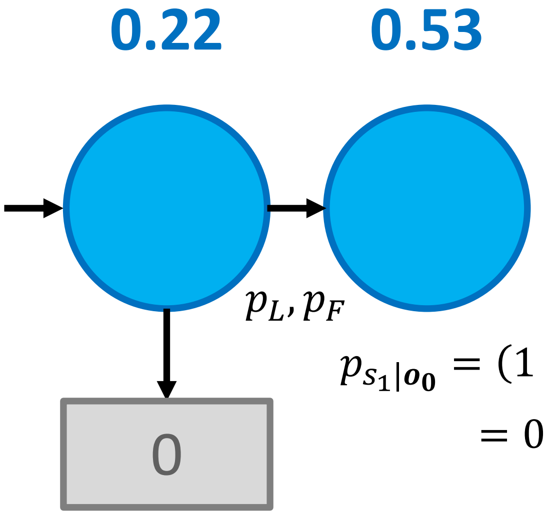
$p_0 = 0.5$

$p_S = 0.2$

$p_G = 0.3$

$p_L = 0.4$

$p_F = 0.0$



$$p_{s_1|\boldsymbol{o_0}} = (1 - p_F) \cdot p_{s_0|\boldsymbol{o_0}} + p_L \cdot (1 - p_{s_0|\boldsymbol{o_0}})$$
$$= 0.22 + 0.4 \cdot 0.78 = 0.53$$

Making predictions using a BKT model

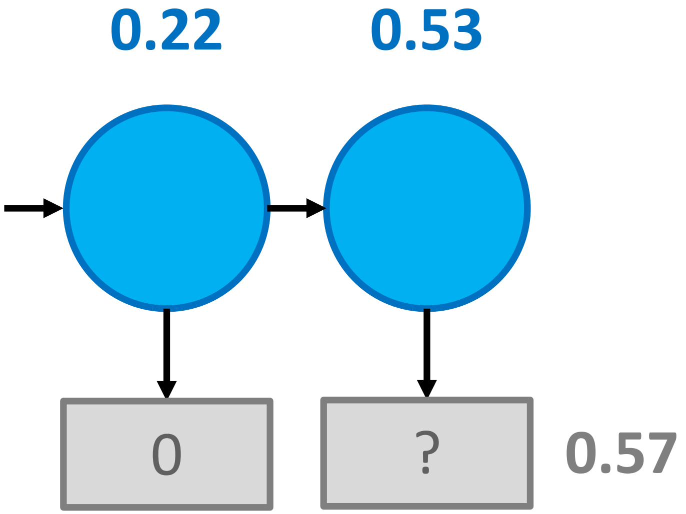
$p_0 = 0.5$

$p_S = 0.2$

$p_G = 0.3$

$p_L = 0.4$

$p_F = 0.0$



$$p(o_1 = 1|\boldsymbol{o_0}) = (1 - p_S) \cdot p_{s_1|\boldsymbol{o_0}} + p_G \cdot (1 - p_{s_1|\boldsymbol{o_0}})$$
$$= 0.8 \cdot 0.53 + 0.3 \cdot 0.47 = 0.57$$

Making predictions using a BKT model

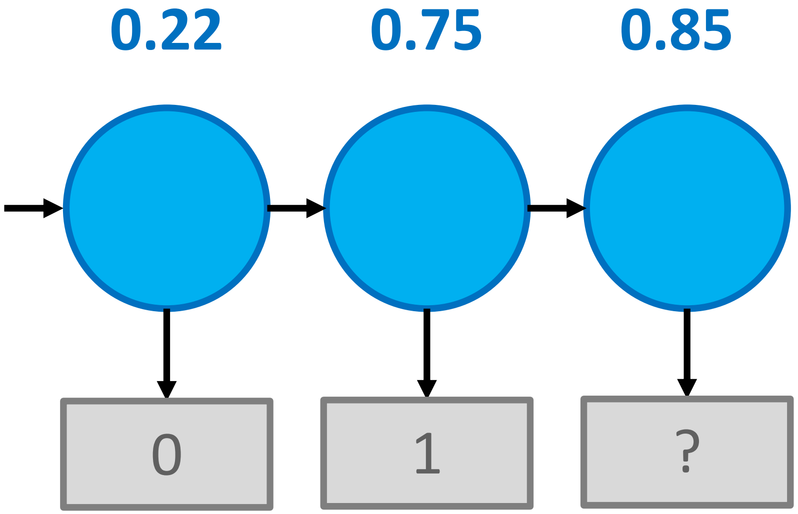
$p_0 = 0.5$

$p_S = 0.2$

$p_G = 0.3$

$p_L = 0.4$

$p_F = 0.0$



Making predictions using a BKT model

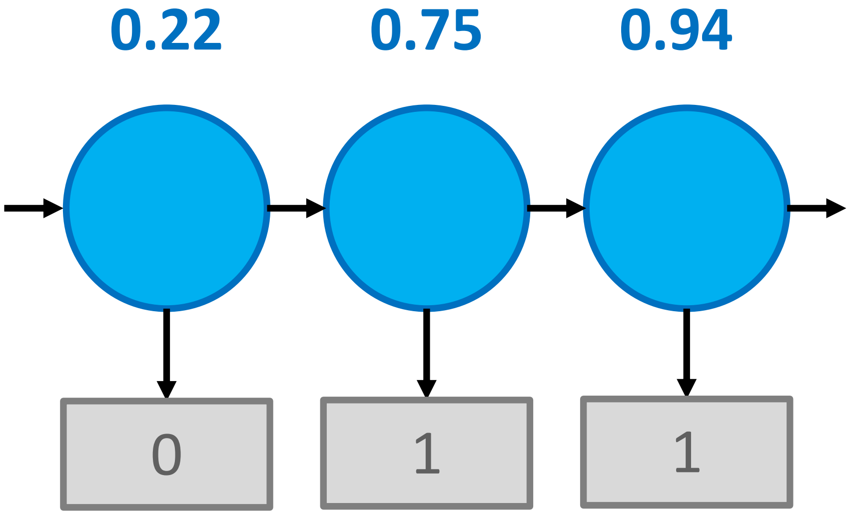
$p_0 = 0.5$

$p_S = 0.2$

$p_G = 0.3$

$p_L = 0.4$

$p_F = 0.0$



Making predictions using a BKT model

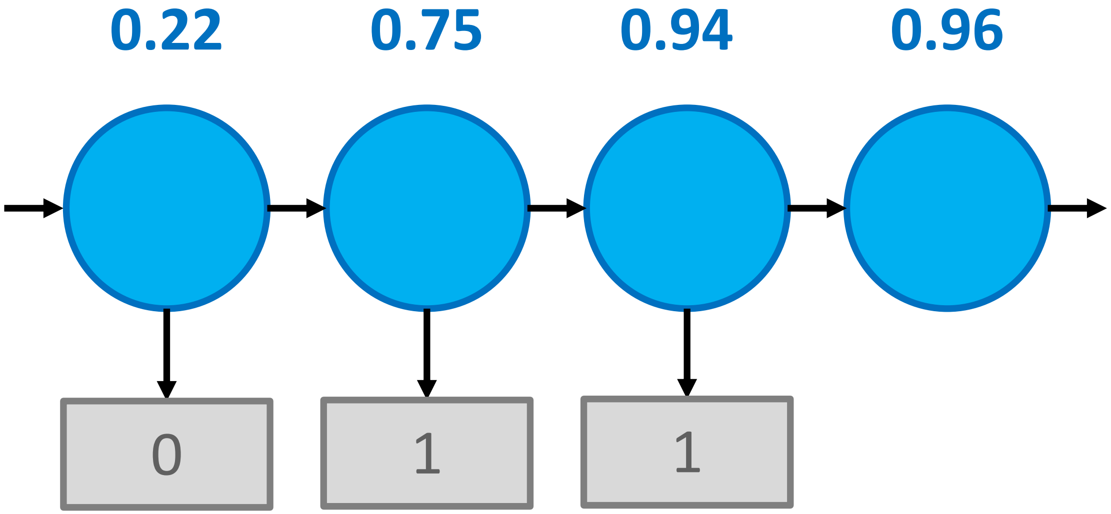
$p_0 = 0.5$

$p_S = 0.2$

$p_G = 0.3$

$p_L = 0.4$

$p_F = 0.0$



Making predictions using a BKT model

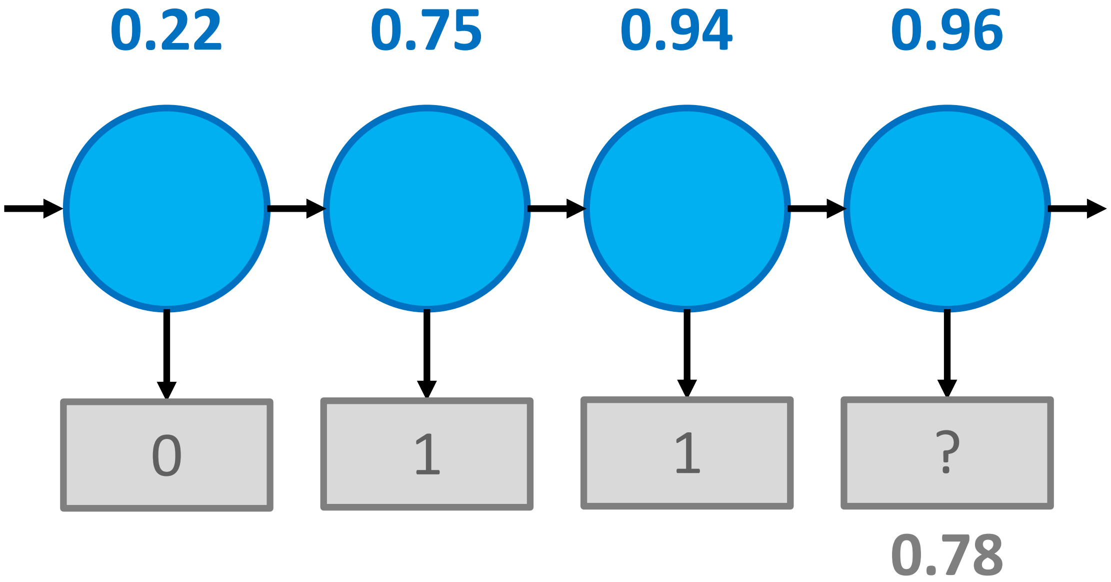
$p_0 = 0.5$

$p_S = 0.2$

$p_G = 0.3$

$p_L = 0.4$

$p_F = 0.0$



Inference in BKT models

$\boldsymbol{o_{t-1}} = [o_0, \ldots, o_{t-1}]$

Equations for time $t = 0$:

Belief about latent state before observation

$$p(s_0 = 1) = p_0$$

Predicted observation at time t

$$p(o_0 = 1) = (1 - p_S) \cdot p_0 + p_G \cdot (1 - p_0)$$
$$p(o_0 = 0) = p_S \cdot p_0 + (1 - p_G) \cdot (1 - p_0)$$

Posterior: belief about latent state after observation

$$\left.\begin{aligned} p_{s_0|1} &= \frac{(1 - p_S) \cdot p_0}{(1 - p_S) \cdot p_0 + p_G \cdot (1 - p_0)} \\ p_{s_0|0} &= \frac{p_S \cdot p_0}{p_S \cdot p_0 + (1 - p_G) \cdot (1 - p_0)} \end{aligned}\right\} p_{s_0|o_0}$$

Equations for time steps $t = 1, \ldots, T$:

$$p_{s_t|\boldsymbol{o_{t-1}}} = (1 - p_F) \cdot p_{s_{t-1}|\boldsymbol{o_{t-1}}} + p_L \cdot (1 - p_{s_{t-1}|\boldsymbol{o_{t-1}}})$$

$$p(o_t = 1|\boldsymbol{o_{t-1}}) = (1 - p_S) \cdot p_{s_t|\boldsymbol{o_{t-1}}} + p_G \cdot (1 - p_{s_t|\boldsymbol{o_{t-1}}})$$
$$p(o_t = 0|\boldsymbol{o_{t-1}}) = p_S \cdot p_{s_t|\boldsymbol{o_{t-1}}} + (1 - p_G) \cdot (1 - p_{s_t|\boldsymbol{o_{t-1}}})$$

$$\left.\begin{aligned} p_{s_t|1,\boldsymbol{o_{t-1}}} &= \frac{(1 - p_S) \cdot p_{s_t|\boldsymbol{o_{t-1}}}}{(1 - p_S) \cdot p_{s_t|\boldsymbol{o_{t-1}}} + p_G \cdot (1 - p_{s_t|\boldsymbol{o_{t-1}}})} \\ p_{s_t|0,\boldsymbol{o_{t-1}}} &= \frac{p_S \cdot p_{s_t|\boldsymbol{o_{t-1}}}}{p_S \cdot p_{s_t|\boldsymbol{o_{t-1}}} + (1 - p_G) \cdot (1 - p_{s_t|\boldsymbol{o_{t-1}}})} \end{aligned}\right\} p_{s_t|\boldsymbol{o_t}}$$

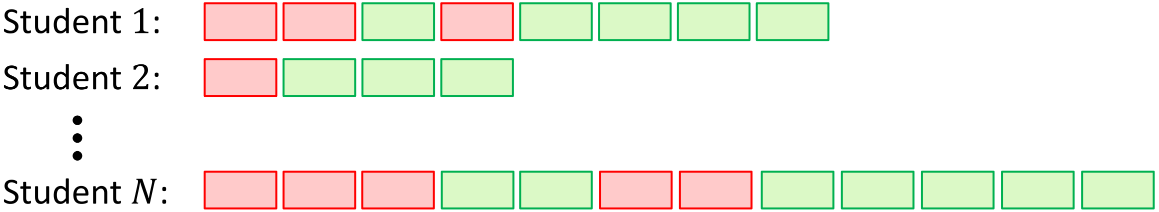
[Corbett & Anderson, UMUAI 1994]

Two tasks need to be solved in practice

- Given a model with parameters $\theta = \{p_0, p_L, p_F, p_S, p_G\}$ and a sequence of observations $\boldsymbol{o} = [o_0, \ldots, o_t]$ from a student s, predict o_{t+1}

Inference

- Given sequences of observations $\boldsymbol{o} = [o_0, \ldots, o_T]$ of N students, learn the parameters $\theta = \{p_0, p_L, p_F, p_S, p_G\}$ that maximize the likelihood of the observed data



Parameter Learning

Tracing Knowledge – why is it useful?

- Is the student learning?
  - Measure what the student *knows* at a specific time *t*
  - More specifically: knowledge of the student about relevant knowledge components (skills)

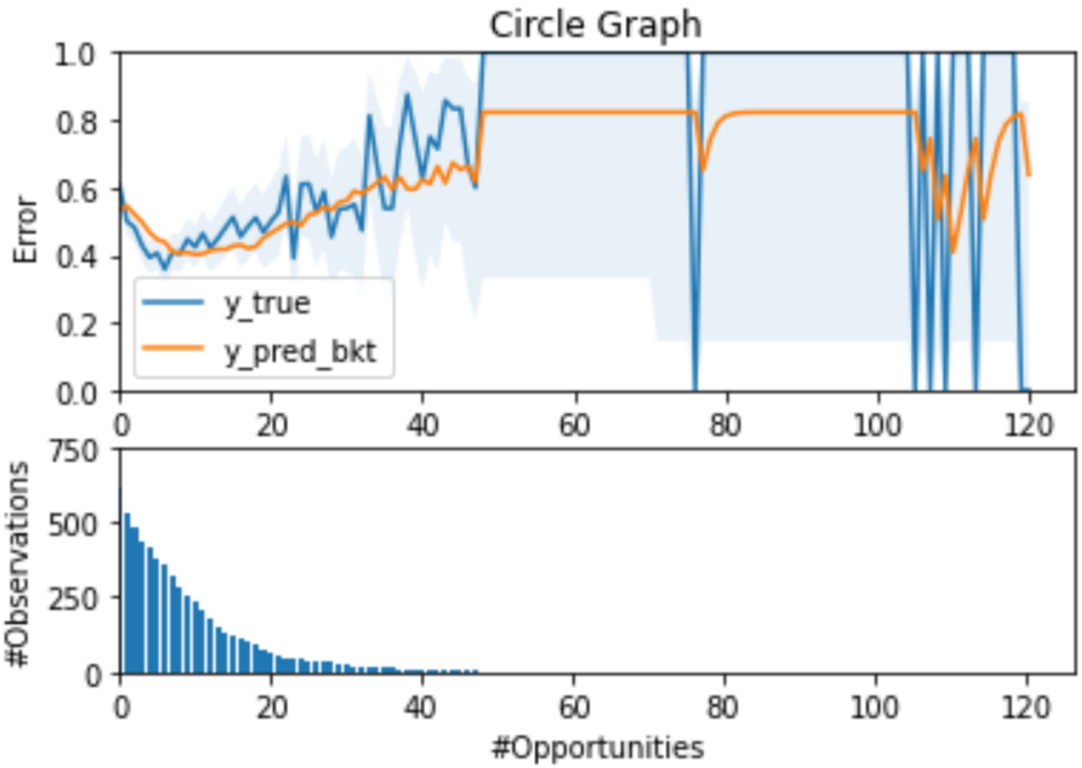
➡ Choose the next appropriate activity

➡ Know which activities support learning

Building a learning curve for skill s

Student	Opportunity	y_true	y_pred
0	0	0	0.3
0	1	0	0.5
0	2	1	0.7
0	3	1	0.9
1	0	0	0.3
1	1	1	0.5
2	0	0	0.3
2	1	1	0.5
2	2	1	0.7
3	0	1	0.3
3	1	0	0.7
3	2	1	0.5
3	3	1	0.9

What could this curve indicate?



Your Turn

- In the student notebook, you have:
  - A trained BKT model for six selected skills
  - A data frame containing the predictions of the BKT model for each observation in the test set
- Your task:
  - Compute the RMSE or AUC separately for the two specified skills
  - Generate and interpret the learning curves for specific skills

Summary – Knowledge Tracing

- Bayesian Knowledge Tracing and Learning Curves
- There are lots of other modeling approaches
  - Factor-based models (AFM & PFA)
  - Deep learning models (DKT)

➡ We will provide links to lecture recordings, notebooks, and exercises on our GitHub