

# Knowledge Tracing

Machine Learning for Behavioral Data

March 27, 2023

# Today's Topic

Week	Lecture/Lab
1	Introduction
2	Data Exploration
3	Regression
4	Classification
5	Model Evaluation
6	<b>Knowledge Tracing</b>
7	Knowledge Tracing
8	Spring Break

Supervised learning on time series:

- Probabilistic graphical models
- Neural networks: LSTM, GRU, etc.

# Getting ready for today's lecture...

- **If not done yet:** clone the repository containing the Jupyter notebook and data for today's lecture into your Noto workspace.
- SpeakUp room for today's lecture:

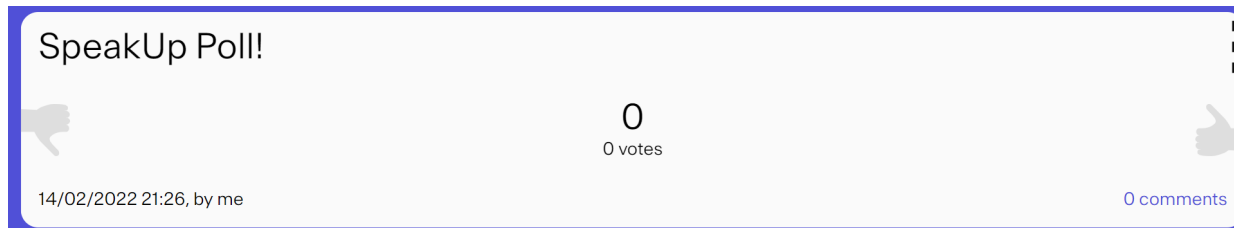
<https://go.epfl.ch/speakup-mlbd>



# Short quiz about the past...

[Model Evaluation] Given a data set  $\{1,2,3,4\}$  , one possible bootstrap set is  $\{1,1,1,1\}$ :

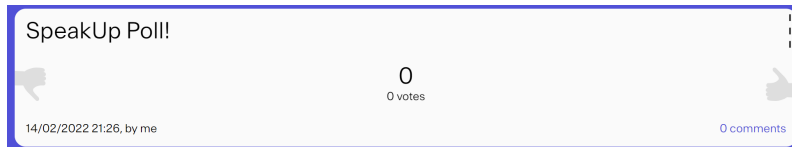
- a) True
- b) False



# Short quiz about the past...

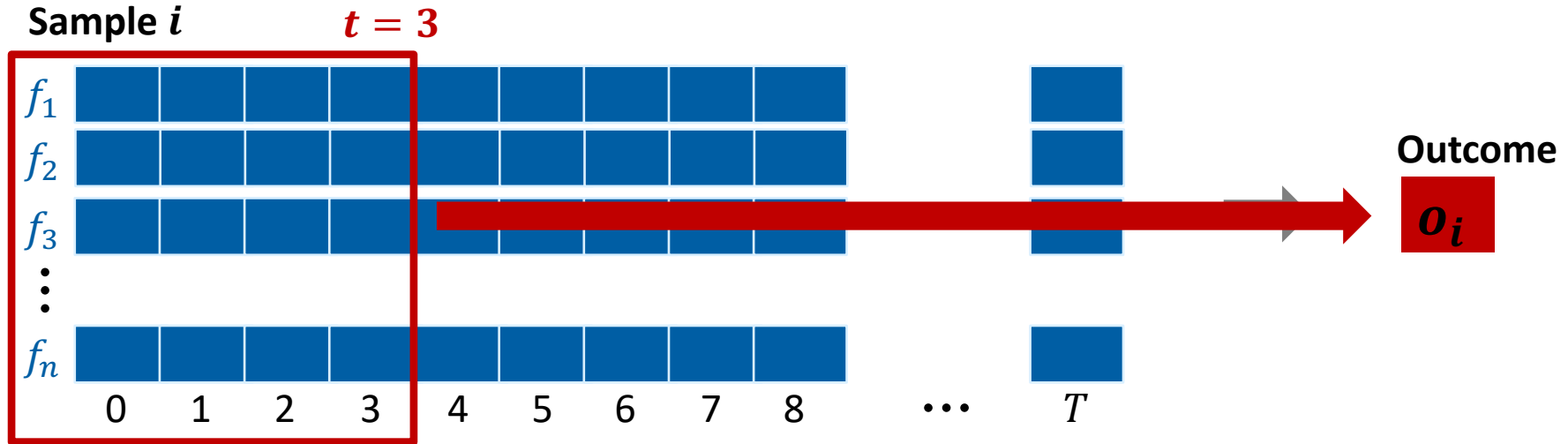
[Model Evaluation] Which of the following statements about  $k$ -fold cross validation are wrong?  $N$  denotes the number of samples in the data set,  $k$  the number of folds:

- a)  $k$  must always be smaller than  $N$ .
- b) The smaller  $k$  is, the more expensive it is to compute the error.
- c) Cross validation can be used to tune model hyperparameters.
- d) Cross validation is not a valid method for computing the generalization error of a model.



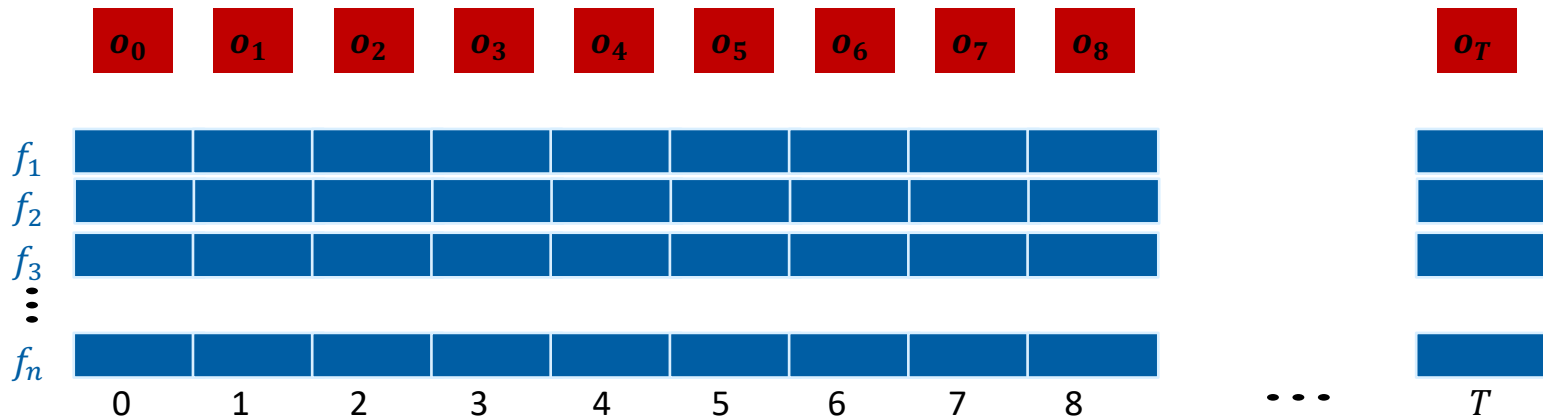
# Time Series – Prediction Task

- Prediction of a target variable after  $t < T$  time steps, where  $T$  is the total number of time steps



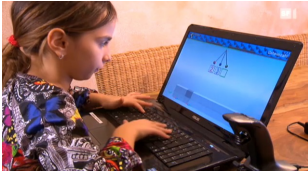
# Time Series – Tracing Task

- Prediction of a target variable after  $t < T$  time steps, where  $T$  is the total number of time steps
- Prediction of a variable in time step  $t + 1$ , based on time steps  $0, \dots, t$



# Today: 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
-

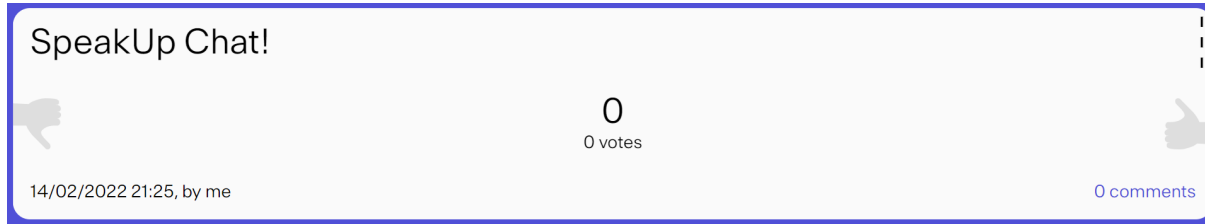
# Today's Use Case

- ASSISTments is a free tool for assigning and assessing math problems and homework
  - All math problems (tasks/items) are associated to a specific skill/knowledge component
  - 4,217 middle-school students
  - 525,534 observations
-

# Today: Tracing Student Knowledge

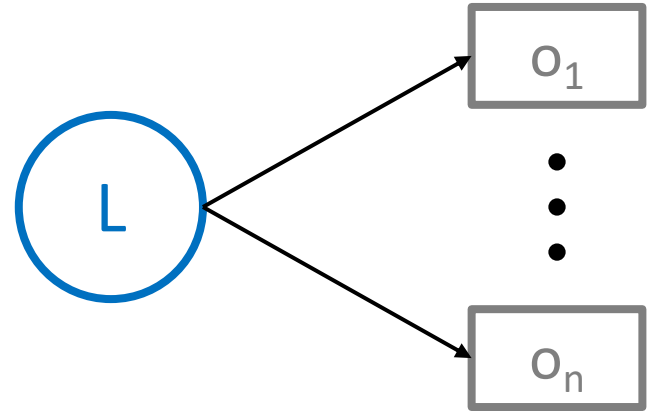
- **Bayesian Knowledge Tracing (BKT)**
    - Latent variables
    - BKT – Inference
    - Practical Example
-

# What is a latent variable?



# What is a latent variable?

- A **latent** variable  $L$  is a variable which is not directly observable/cannot be measured
- It is assumed to affect the outcome of other variables  $\mathbf{o}$ , which can be **observed** (directly measured)

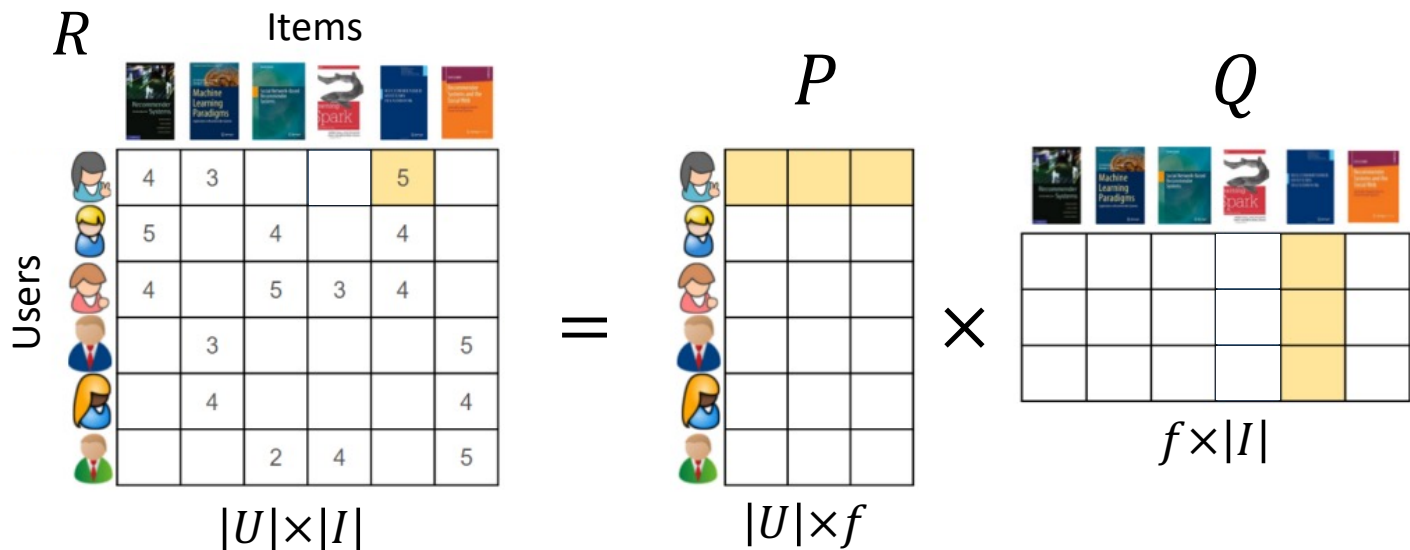


# Why should we use latent variables?

- In many scientific fields, we are interested in concepts/factors that cannot directly be measured/observed:
    - Political sciences: leadership, political competence, etc.
    - Psychology: stress, self-worth, personality characteristics, talent, etc.
    - Education: memory, spatial ability, cognitive abilities, etc.
  - We represent underlying concepts/factors by latent variables and infer them from the observed variables
-

# Example 1: Recommender Systems

- Given: ratings of users  $u$  for items  $i$  (e.g., books)



## Example 2: Education

- Observations: binary answers (correct/wrong) of students to items (tasks)



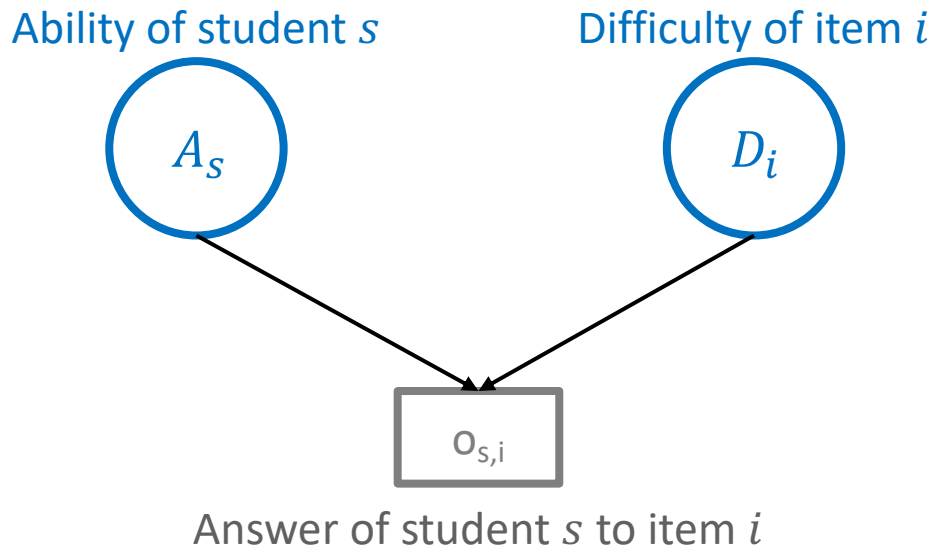
Answer of student  $s$  to item  $i$

---

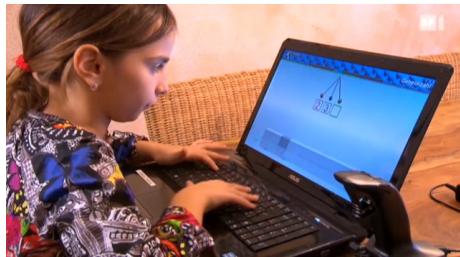


## Example 2: Education

- Observations: binary answers (correct/wrong) of students to items (tasks)



# Is the student learning?

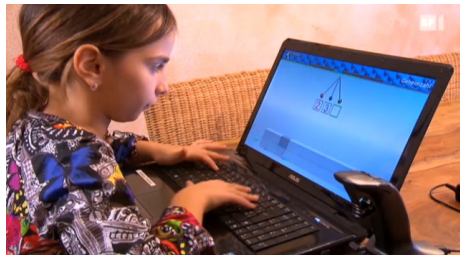


Task:             $50 - 23 = ?$      $75 - 12 = ?$      $38 - 14 = ?$

Answer:            27                      61                      24

---

# What are we measuring?



Task:             $50 - 23 = ?$      $75 - 12 = ?$      $38 - 14 = ?$

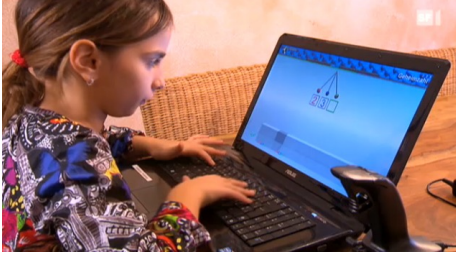
Answer:            27                      61                      24

1

0

1

# Binary observations of student answers



**Subtraction 0-100**

1

2

...

n

0

0

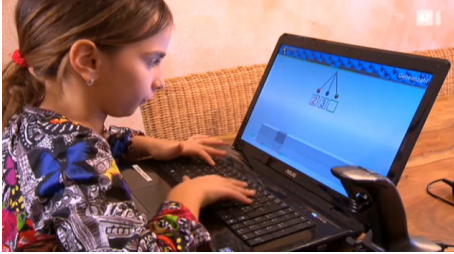
1

0

1

1

# Predicting future performance



**Subtraction 0-100**

1

2

...

n

n+1

0

0

1

0

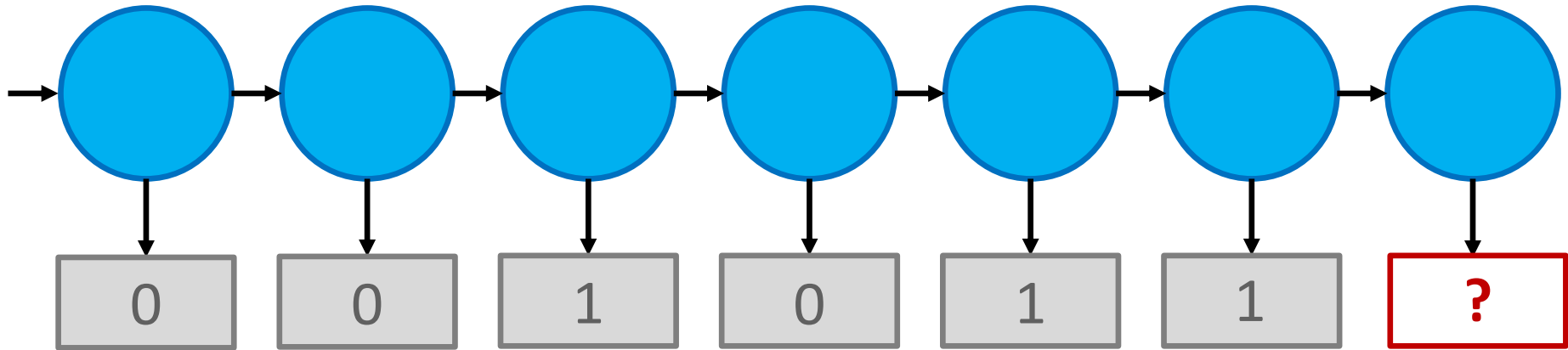
1

1

?

# Bayesian Knowledge Tracing (BKT)

● Latent variable    □ Observed variable

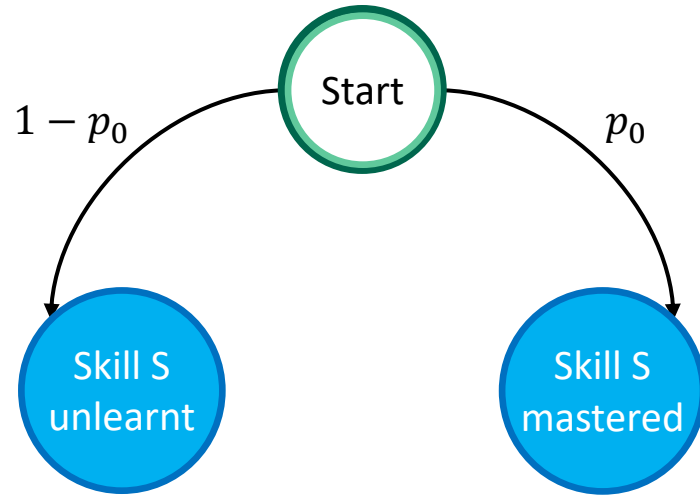


# Bayesian Knowledge Tracing (BKT)



Observations for student  $s$ :

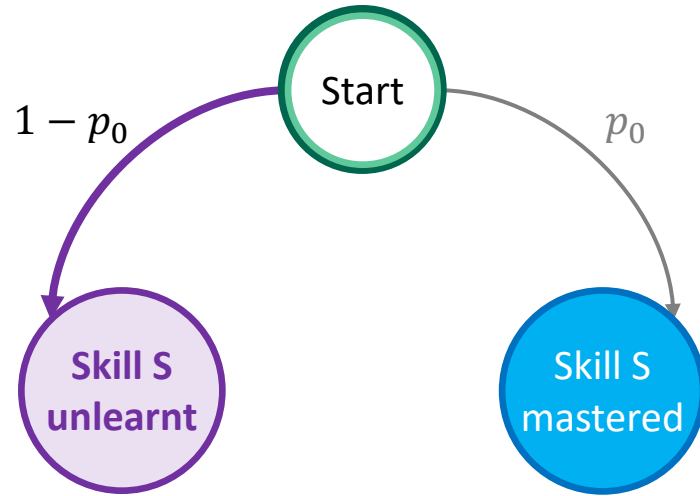
# Bayesian Knowledge Tracing (BKT)



Observations for student  $s$ :

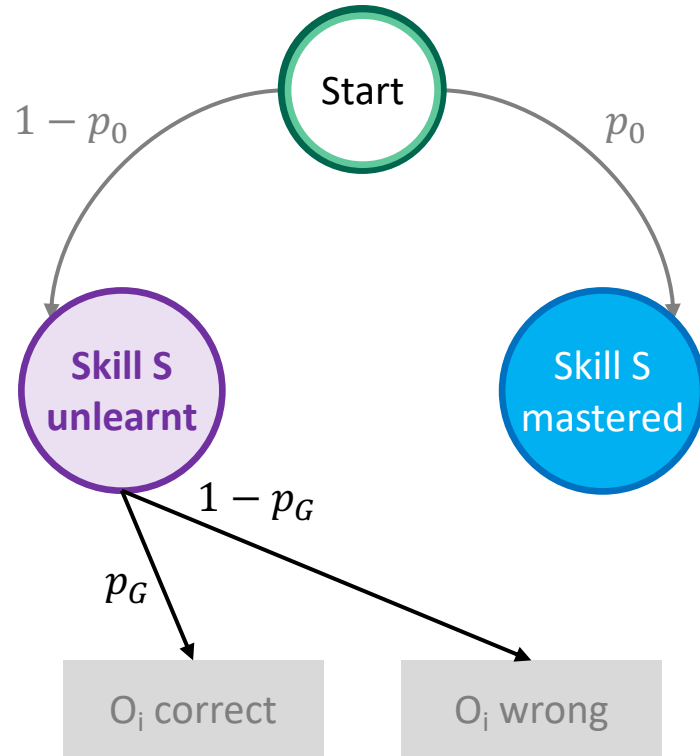


# Bayesian Knowledge Tracing (BKT)



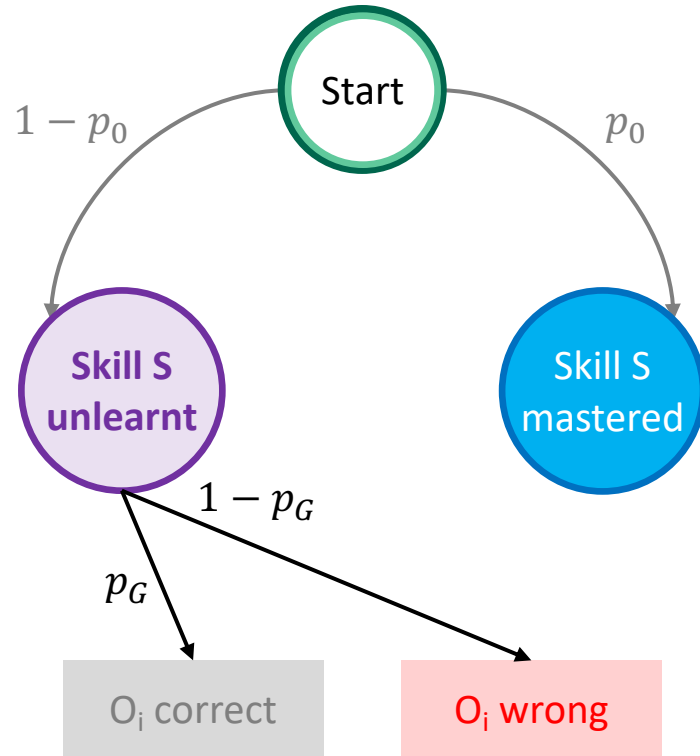
Observations for student  $s$ :  
 $t = 0$ :

# Bayesian Knowledge Tracing (BKT)



Observations for student  $s$ :  
 $t = 0$ :

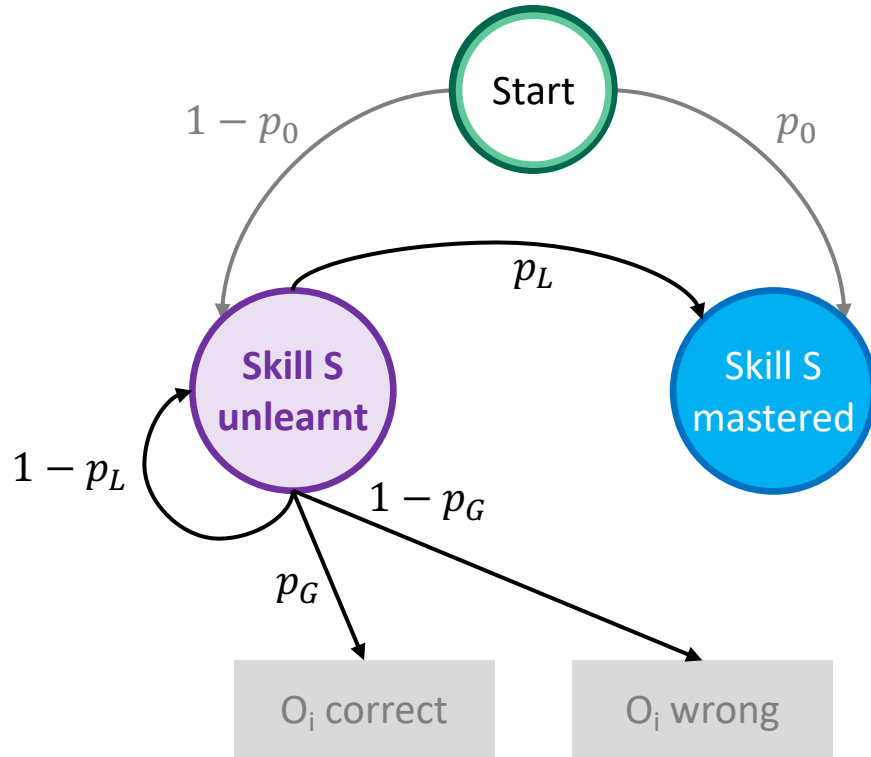
# Bayesian Knowledge Tracing (BKT)



Observations for student  $s$ :

$t = 0$ : **0**

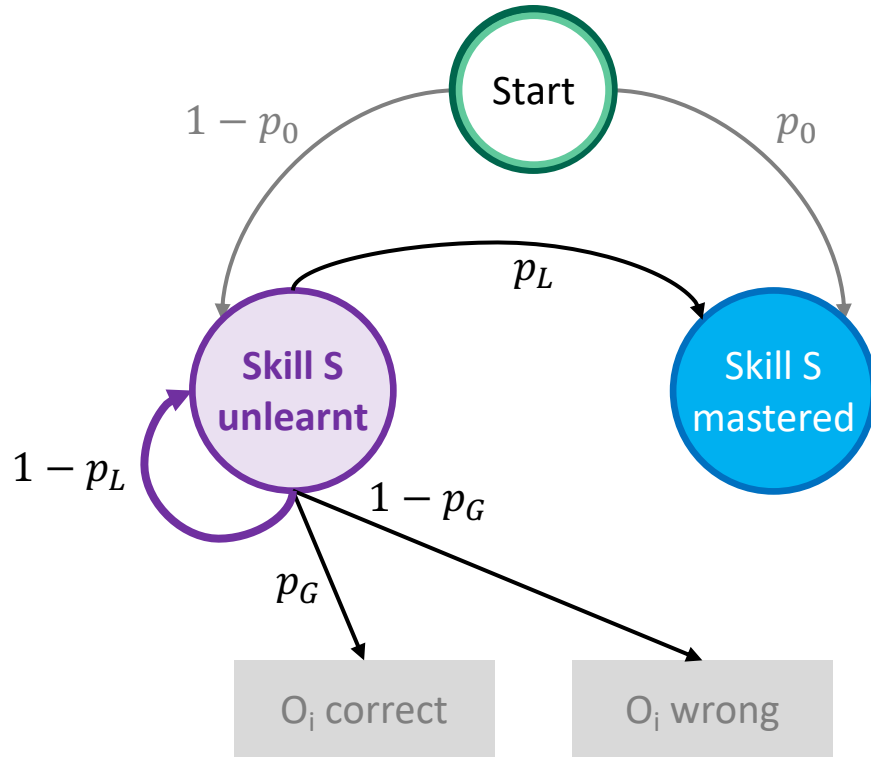
# Bayesian Knowledge Tracing (BKT)



Observations for student  $s$ :

$t = 0$ : 0

# Bayesian Knowledge Tracing (BKT)

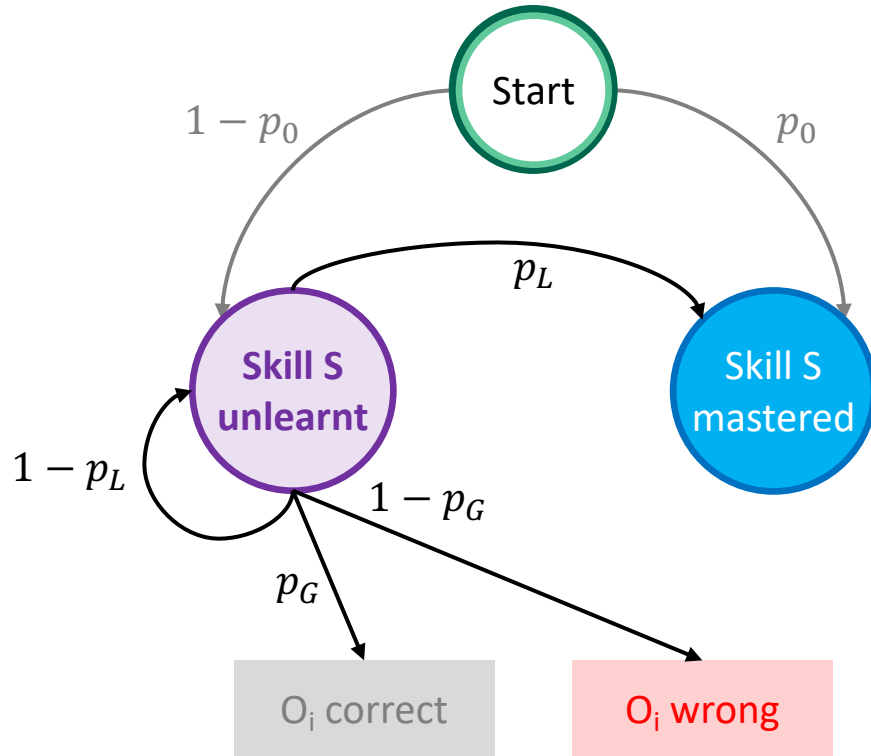


Observations for student  $s$ :

$t = 0$ : 0

$t = 1$ :

# Bayesian Knowledge Tracing (BKT)

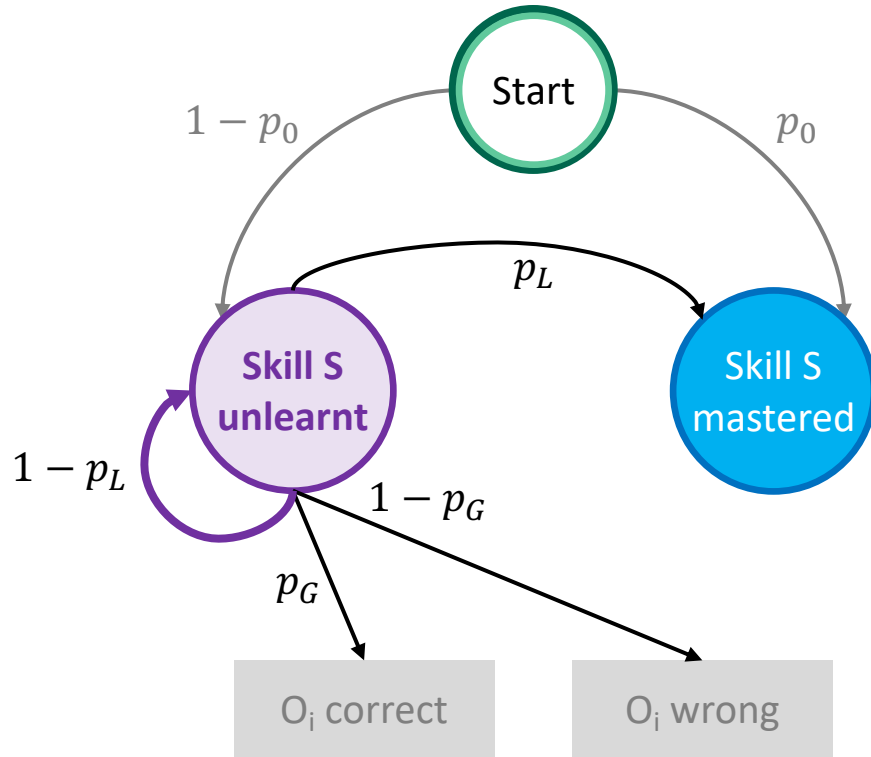


Observations for student  $s$ :

$t = 0$ : 0

$t = 1$ : 0

# Bayesian Knowledge Tracing (BKT)



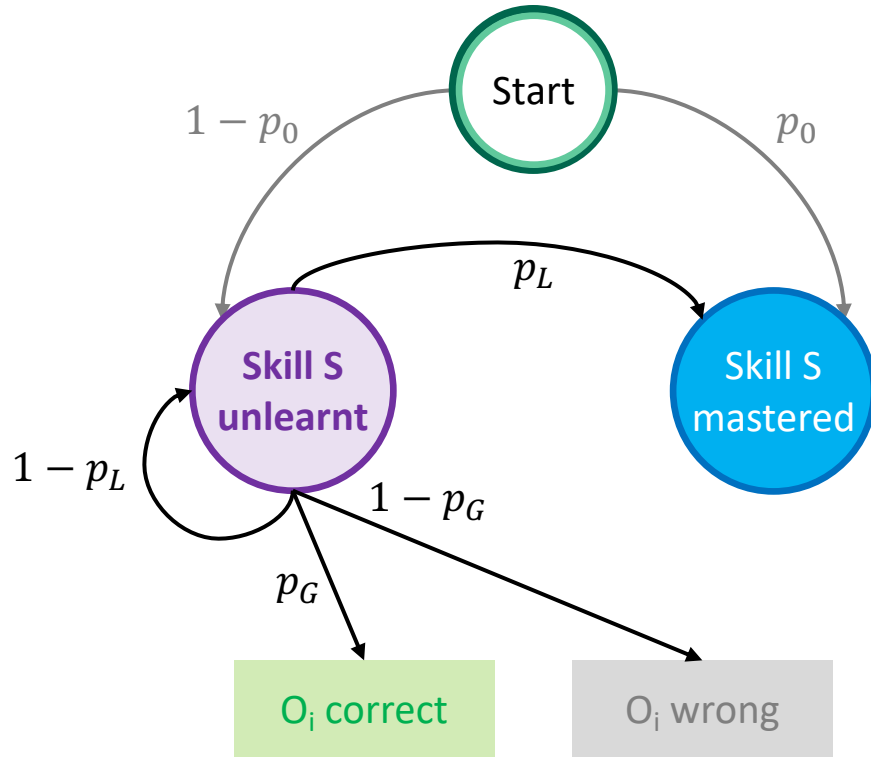
Observations for student  $s$ :

$t = 0$ : 0

$t = 1$ : 0

$t = 2$ :

# Bayesian Knowledge Tracing (BKT)



Observations for student  $s$ :

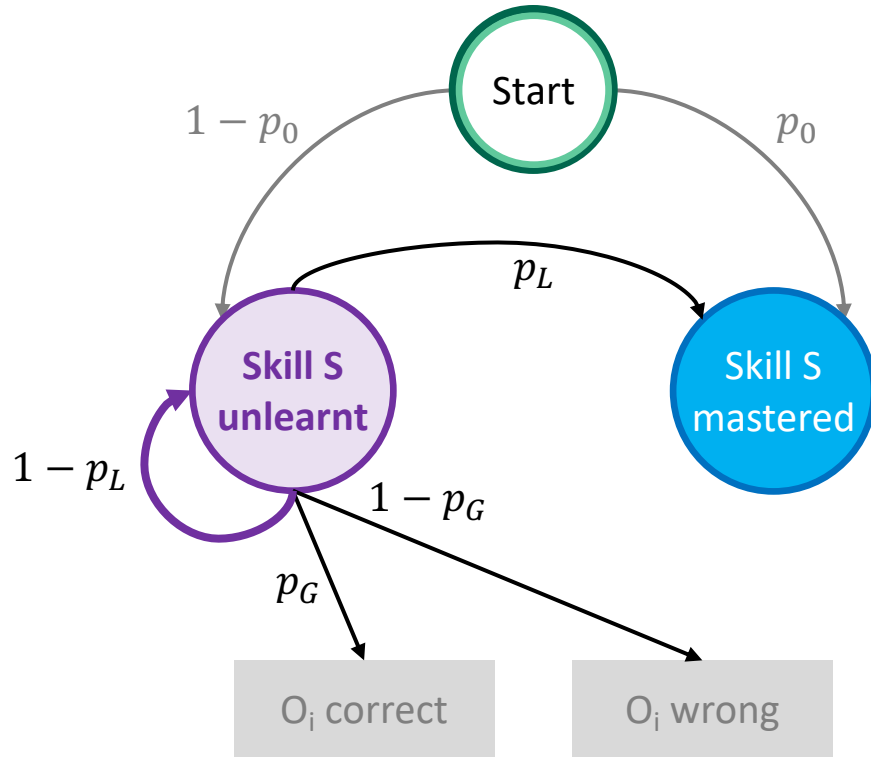
$t = 0$ : 0

$t = 1$ : 0

$t = 2$ : 1



# Bayesian Knowledge Tracing (BKT)



Observations for student  $s$ :

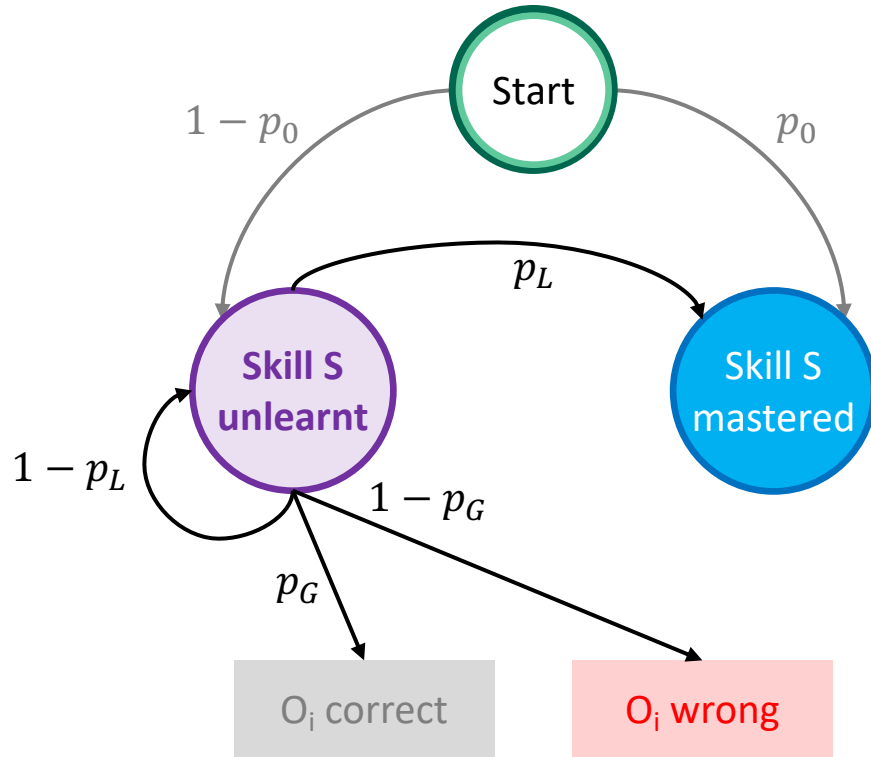
$t = 0$ : 0

$t = 1$ : 0

$t = 2$ : 1

$t = 3$ :

# Bayesian Knowledge Tracing (BKT)



Observations for student  $s$ :

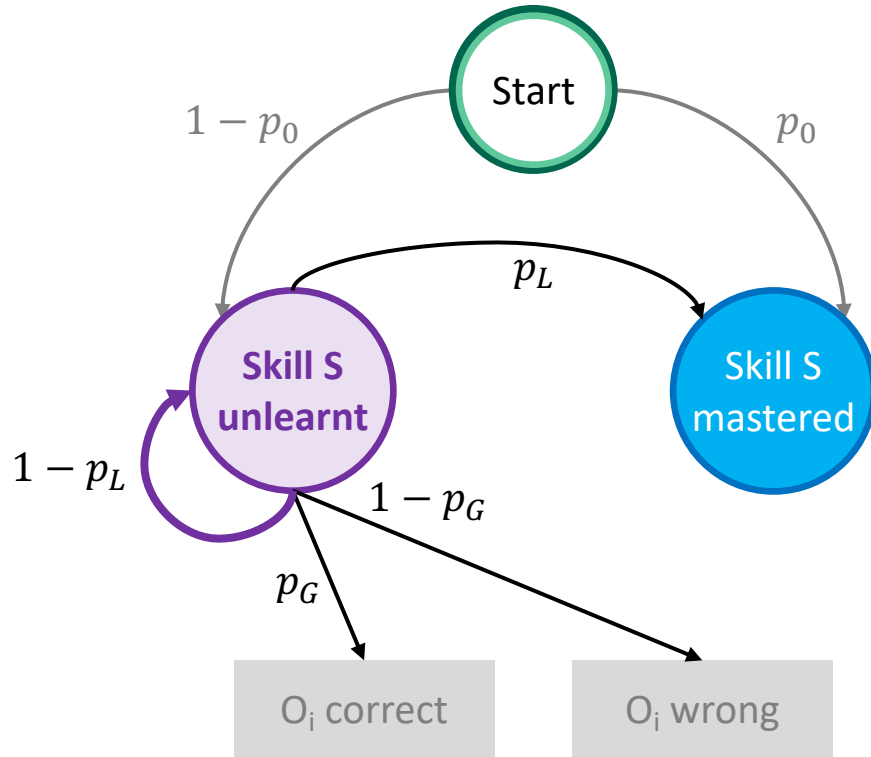
$t = 0$ : 0

$t = 1$ : 0

$t = 2$ : 1

$t = 3$ : 0

# Bayesian Knowledge Tracing (BKT)



Observations for student  $s$ :

$t = 0$ : 0

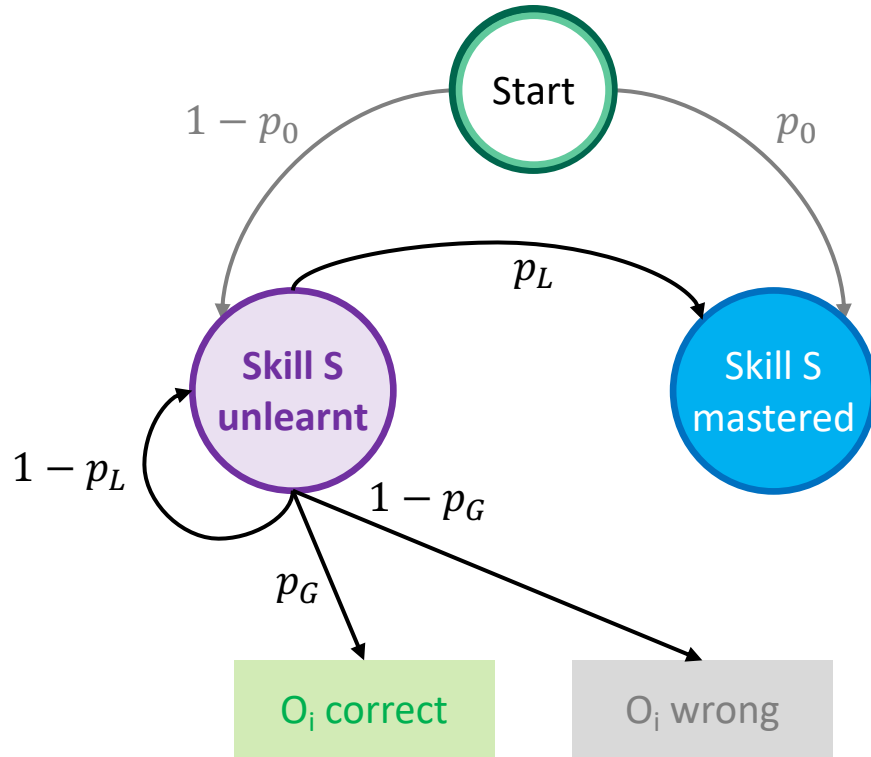
$t = 1$ : 0

$t = 2$ : 1

$t = 3$ : 0

$t = 4$ :

# Bayesian Knowledge Tracing (BKT)



Observations for student  $s$ :

$t = 0$ : 0

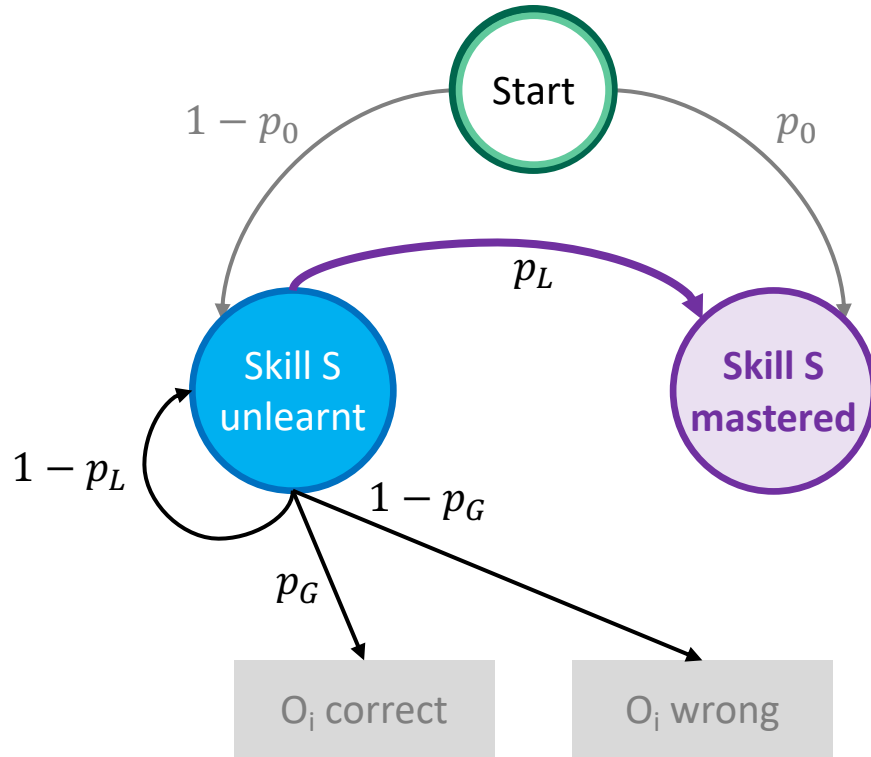
$t = 1$ : 0

$t = 2$ : 1

$t = 3$ : 0

$t = 4$ : 1

# Bayesian Knowledge Tracing (BKT)



Observations for student  $s$ :

$t = 0$ : 0

$t = 1$ : 0

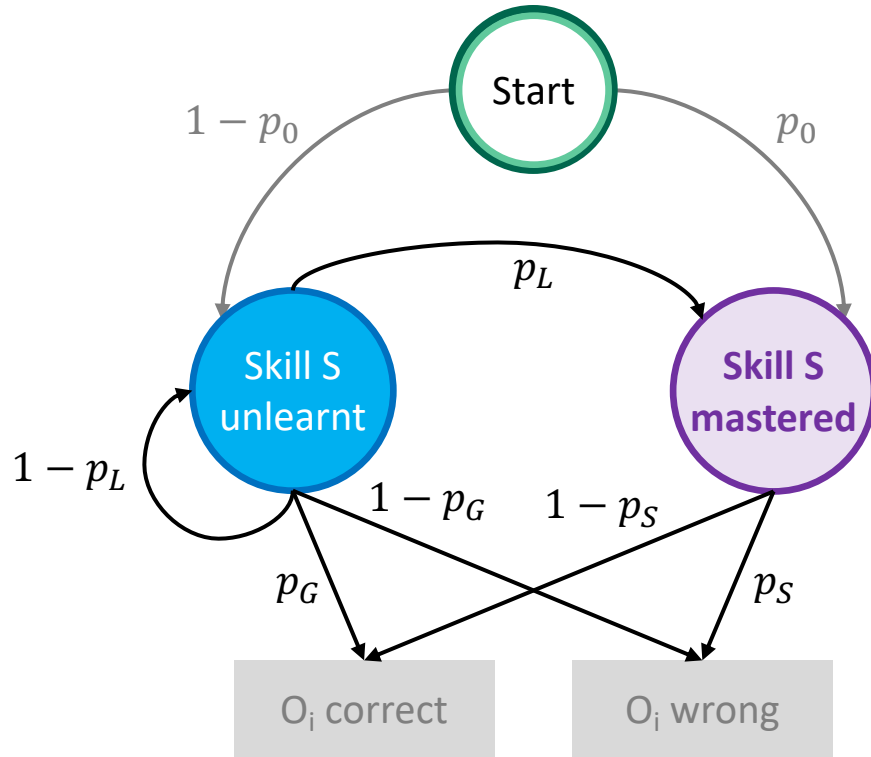
$t = 2$ : 1

$t = 3$ : 0

$t = 4$ : 1

$t = 5$ :

# Bayesian Knowledge Tracing (BKT)



Observations for student  $s$ :

$t = 0$ : 0

$t = 1$ : 0

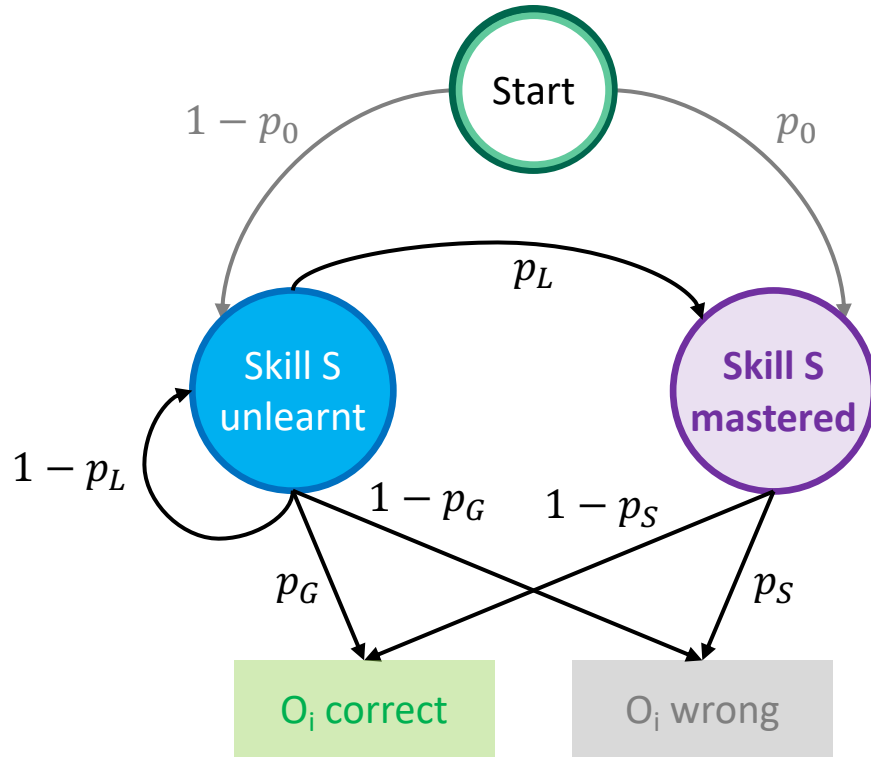
$t = 2$ : 1

$t = 3$ : 0

$t = 4$ : 1

$t = 5$ :

# Bayesian Knowledge Tracing (BKT)



Observations for student  $s$ :

$t = 0$ : 0

$t = 1$ : 0

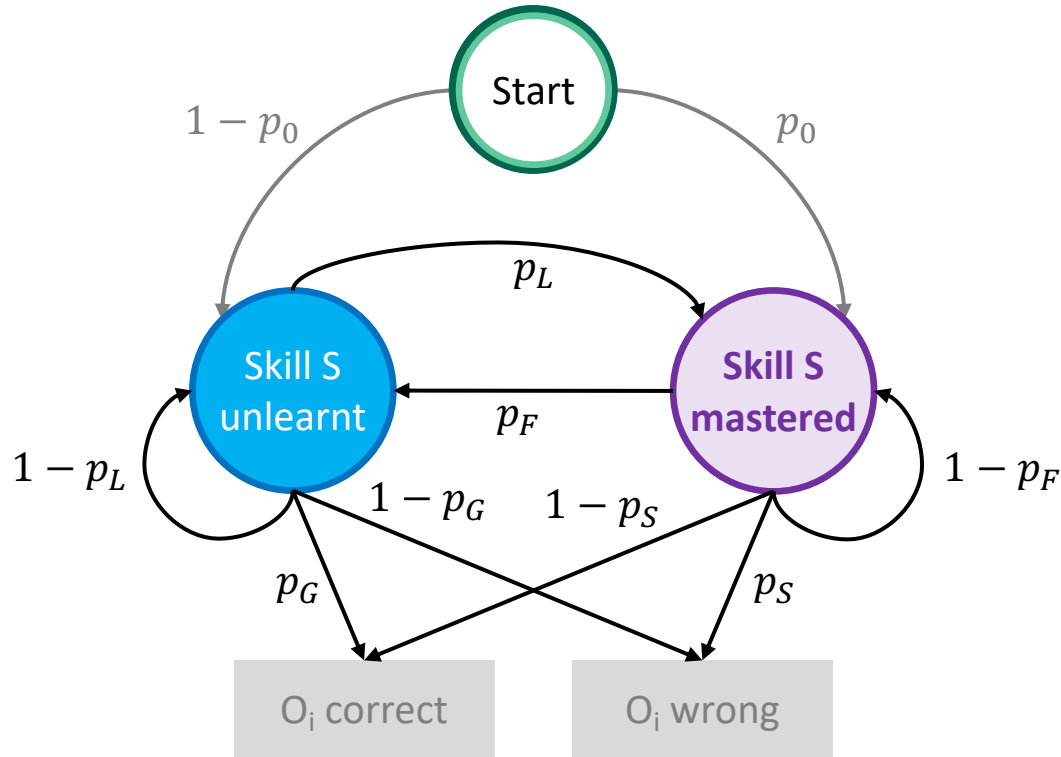
$t = 2$ : 1

$t = 3$ : 0

$t = 4$ : 1

$t = 5$ : 1

# Bayesian Knowledge Tracing (BKT)



Observations for student  $s$ :

$t = 0$ : 0

$t = 1$ : 0

$t = 2$ : 1

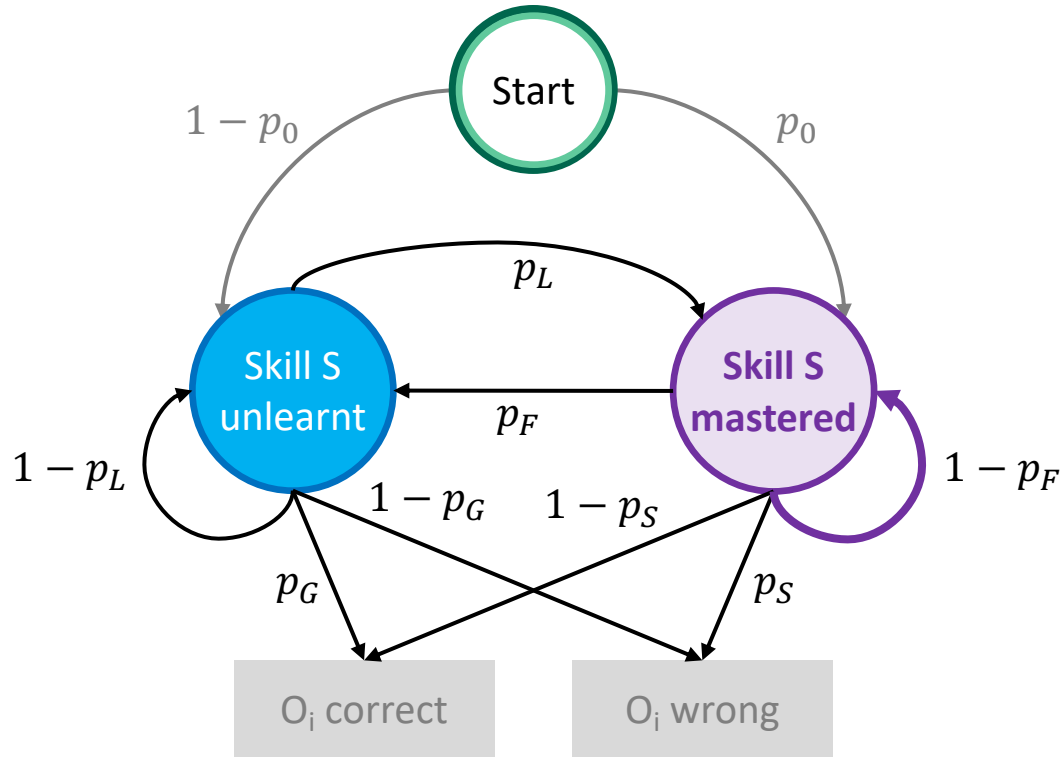
$t = 3$ : 0

$t = 4$ : 1

$t = 5$ : 1



# Bayesian Knowledge Tracing (BKT)



Observations for student  $s$ :

$t = 0$ : 0

$t = 1$ : 0

$t = 2$ : 1

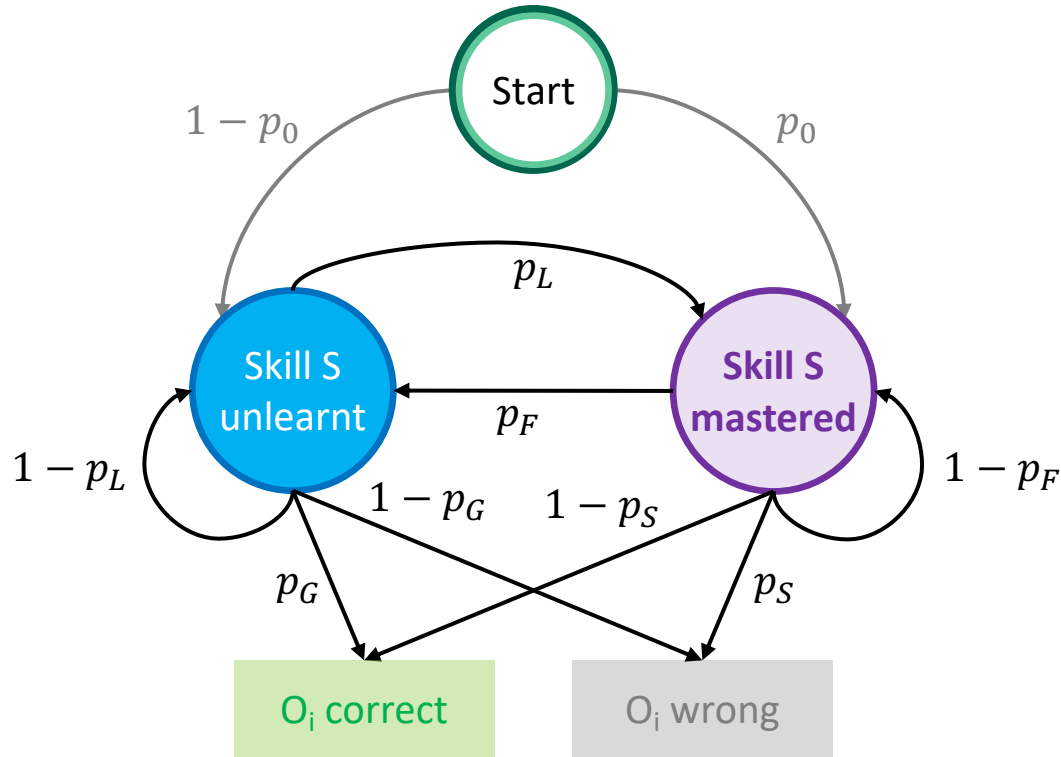
$t = 3$ : 0

$t = 4$ : 1

$t = 5$ : 1

$t = 6$ :

# Bayesian Knowledge Tracing (BKT)



Observations for student  $s$ :

$t = 0$ : 0

$t = 1$ : 0

$t = 2$ : 1

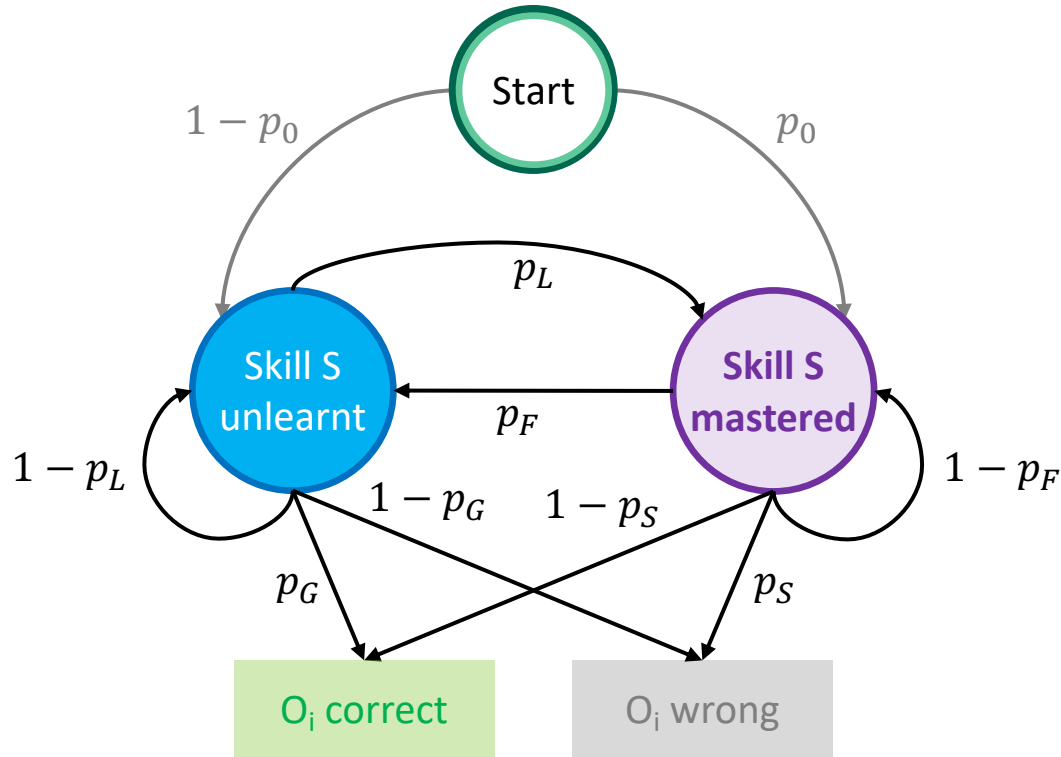
$t = 3$ : 0

$t = 4$ : 1

$t = 5$ : 1

$t = 6$ : 1

# Bayesian Knowledge Tracing (BKT)



Observations for student  $s$ :

$t = 0$ : 0

$t = 1$ : 0

$t = 2$ : 1

$t = 3$ : 0

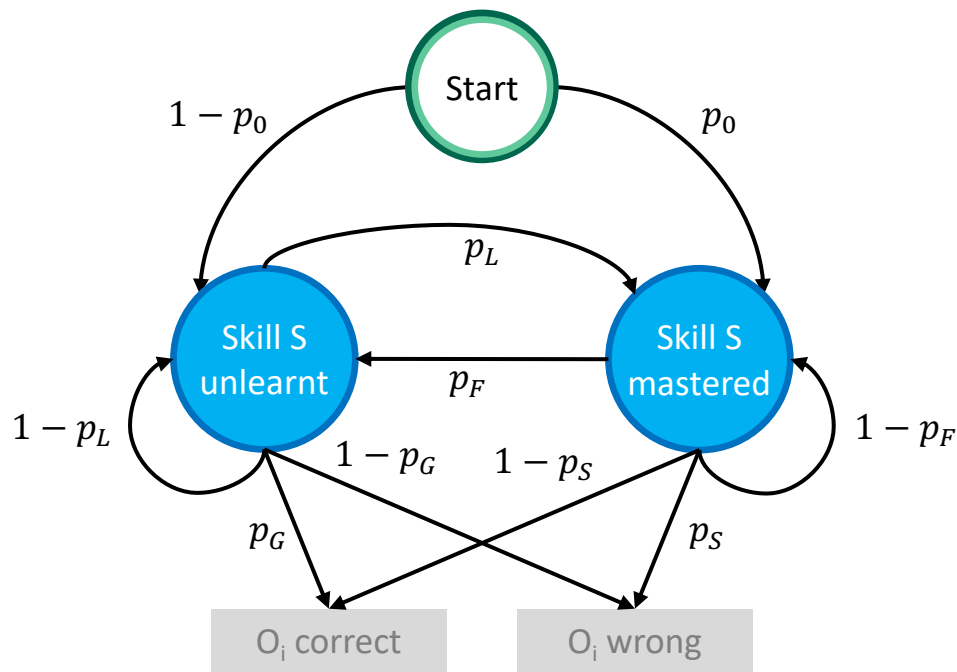
$t = 4$ : 1

$t = 5$ : 1

$t = 6$ : 1

$\mathbf{o}_s = [0, 0, 1, 0, 1, 1, 1]$

# BKT - Terminology

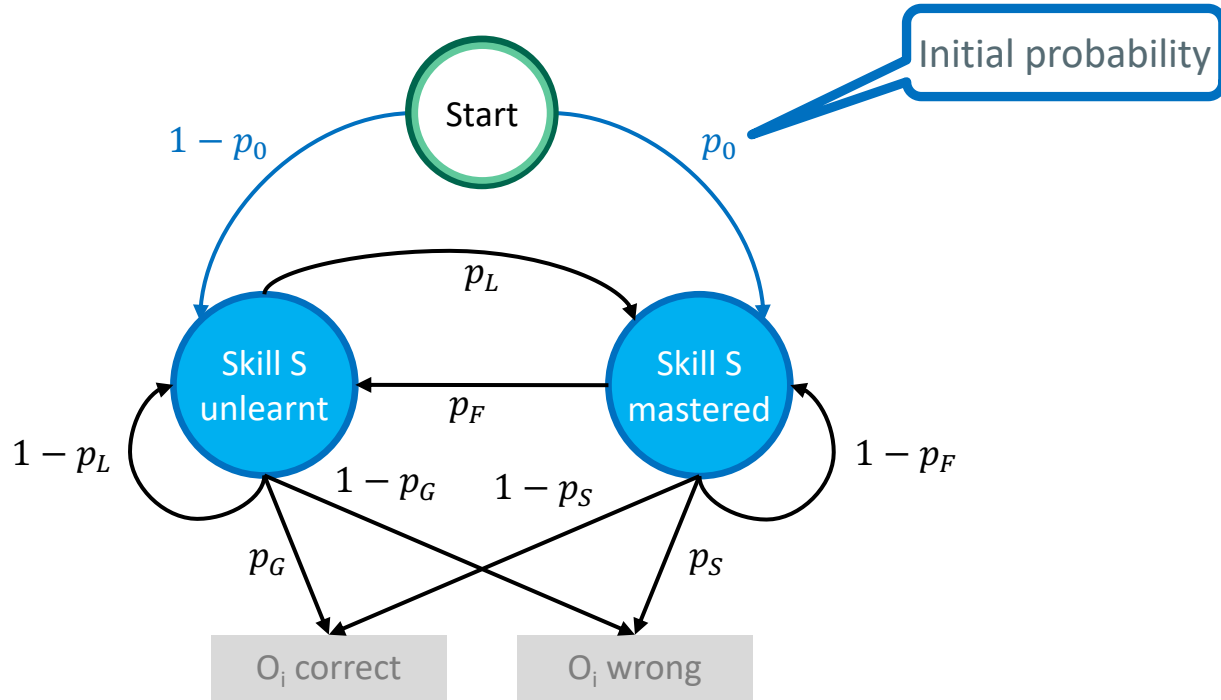


- One **latent** variable ( $S$ ) with two possible states
- Observations (also binary)

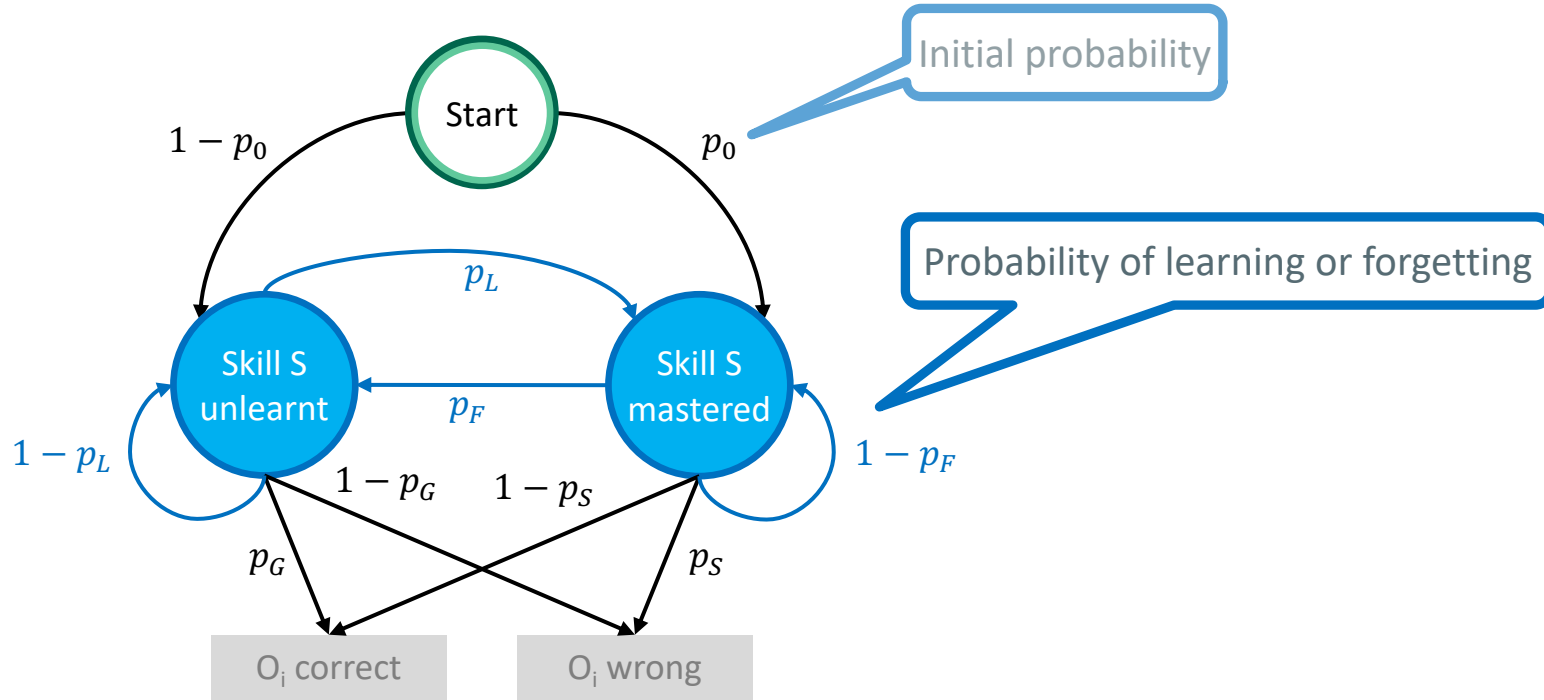
Five parameters:  $\theta = \{p_0, p_L, p_F, p_S, p_G\}$

Transition probabilities

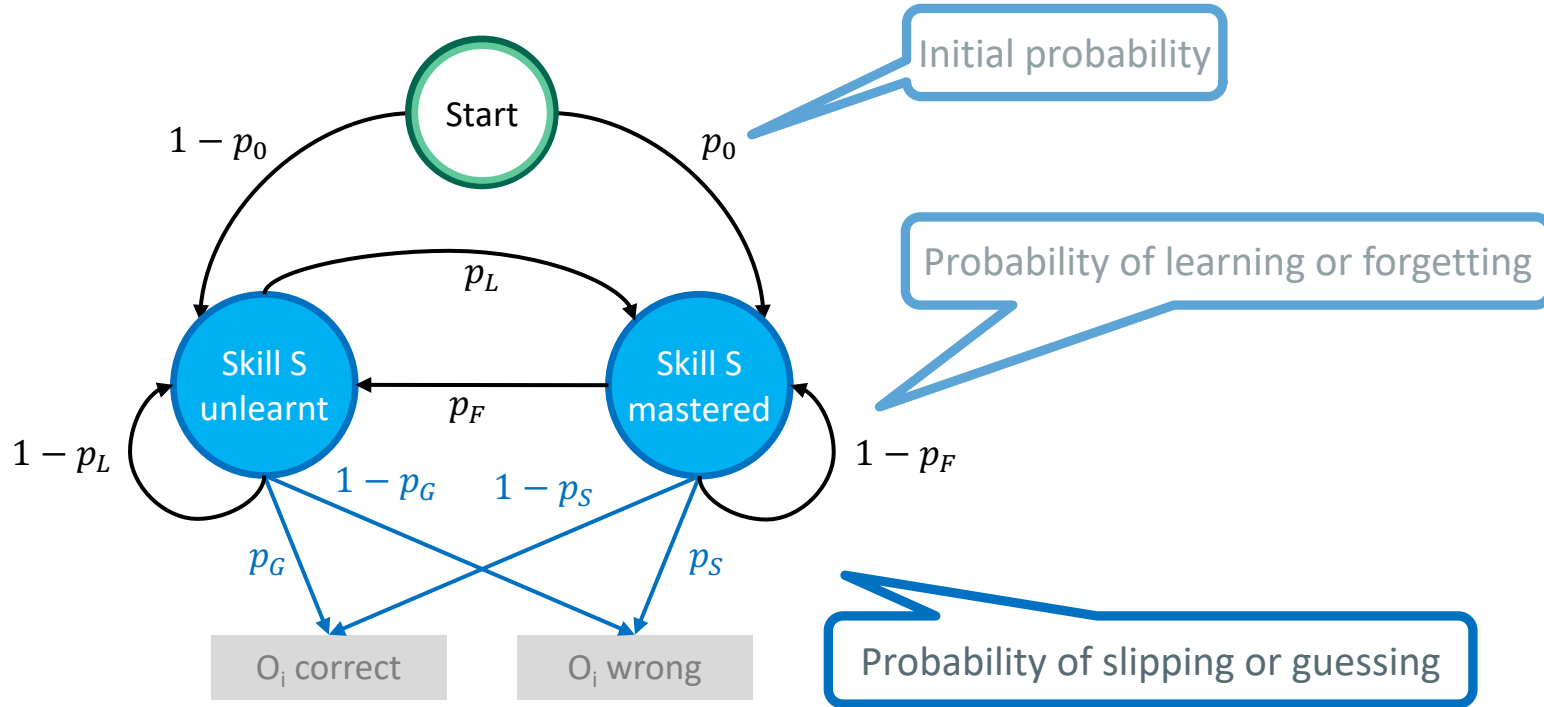
# BKT parameters are interpretable



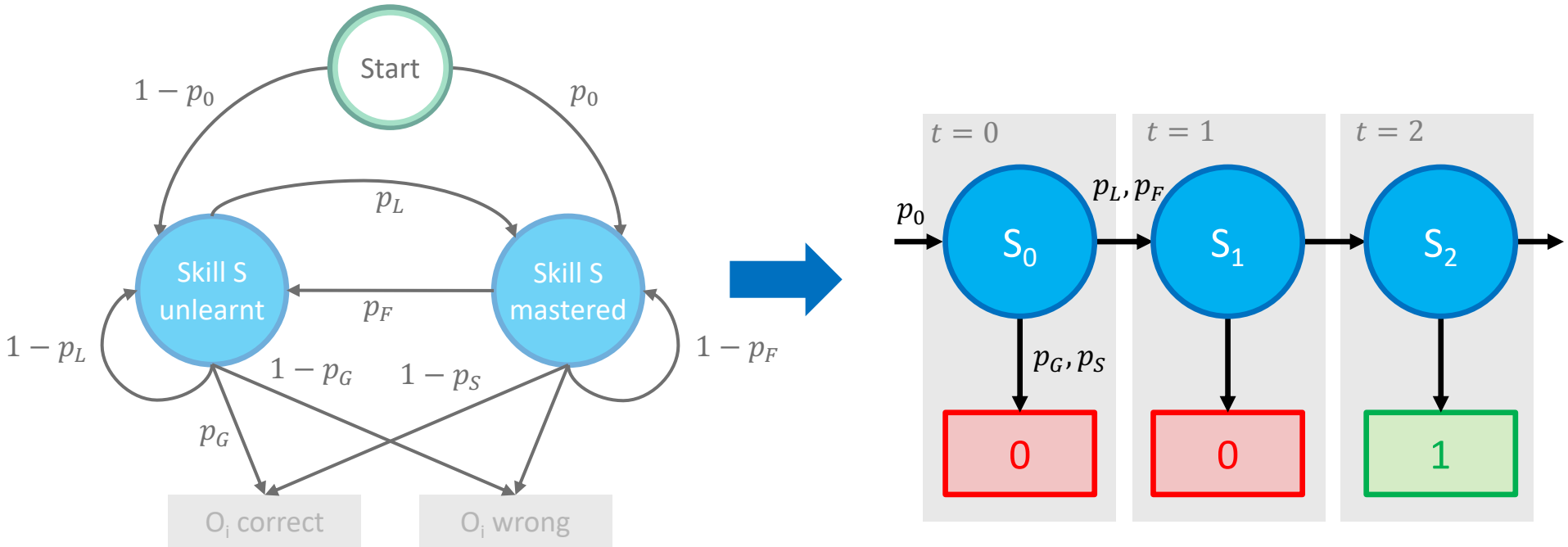
# BKT parameters are interpretable



# BKT parameters are interpretable



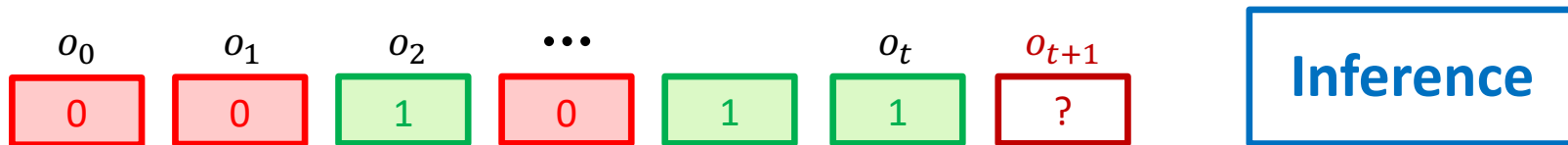
# BKT – unrolled over time



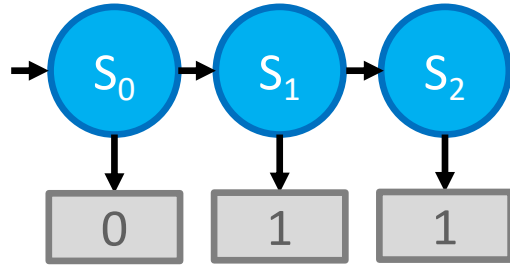


# Two tasks need to be solved in practice

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



# Inference Example



$$p_0 = 0.5$$

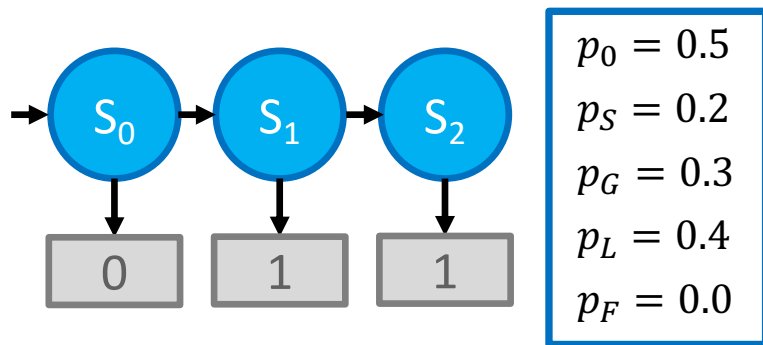
$$p_S = 0.2$$

$$p_G = 0.3$$

$$p_L = 0.4$$

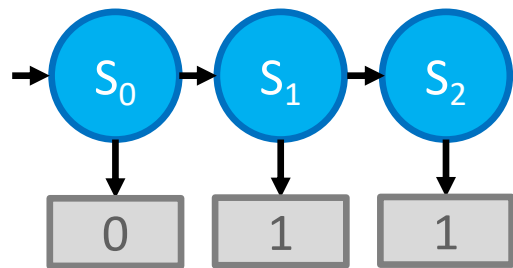
$$p_F = 0.0$$

# Inference Example – Your Turn



- $p(s_0 = 1)$ ?
- $p(o_0 = 1)$ ?
- $p(s_1 = 1 | o_0 = 0)$ ?
- $p(o_1 = 1 | o_0 = 0)$ ?
- $p(s_2 = 1 | o_0 = 0, o_1 = 1)$ ?

# Inference Example – Your Turn



$$\begin{aligned}
 p_0 &= 0.5 \\
 p_S &= 0.2 \\
 p_G &= 0.3 \\
 p_L &= 0.4 \\
 p_F &= 0.0
 \end{aligned}$$

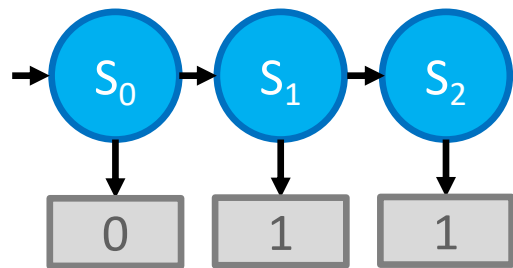
$S_0$	$p(S_0)$
1	$p_0$
0	$1-p_0$

$S_t$	$S_{t+1}$	$p(S_{t+1} S_t)$
0	0	$1 - p_L$
0	1	$p_L$
1	0	$p_F$
1	1	$1 - p_F$

$S_t$	$O_t$	$p(O_t S_t)$
0	0	$1 - p_G$
0	1	$p_G$
1	0	$p_S$
1	1	$1 - p_S$

- $p(s_0 = 1)?$
- $p(o_0 = 1)?$
- $p(s_1 = 1|o_0 = 0)?$
- $p(o_1 = 1|o_0 = 0)?$
- $p(s_2 = 1|o_0 = 0, o_1 = 1)?$

# Inference Example – Your Turn



$$\begin{aligned}
 p_0 &= 0.5 \\
 p_S &= 0.2 \\
 p_G &= 0.3 \\
 p_L &= 0.4 \\
 p_F &= 0.0
 \end{aligned}$$

$S_0$	$p(S_0)$
1	$p_0$
0	$1-p_0$

$S_t$	$S_{t+1}$	$p(S_{t+1} S_t)$
0	0	$1 - p_L$
0	1	$p_L$
1	0	$p_F$
1	1	$1 - p_F$

$S_t$	$O_t$	$p(O_t S_t)$
0	0	$1 - p_G$
0	1	$p_G$
1	0	$p_S$
1	1	$1 - p_S$

Some useful rules:

$$p(A, B) = p(A|B) \cdot p(B)$$

$$p(A|B) = \frac{p(B|A) \cdot p(A)}{p(B)}$$

$$p(A = 1) = p(A = 1, B = 1) + p(A = 1, B = 0)$$

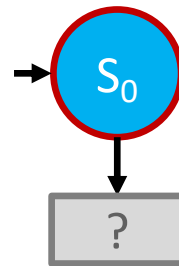
- $p(s_0 = 1)?$
- $p(o_0 = 1)?$
- $p(s_1 = 1|o_0 = 0)?$
- $p(o_1 = 1|o_0 = 0)?$
- $p(s_2 = 1|o_0 = 0, o_1 = 1)?$

# Inference in BKT models

Equations for time step 0:

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

$$p(s_0 = 0) = 1 - p_0$$



# Inference in BKT models

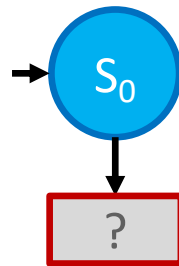
Equations for time step 0:

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

$$p(s_0 = 0) = 1 - p_0$$

$$\begin{aligned} p(o_0 = 1) &= p(o_0 = 1, s_0 = 1) + p(o_0 = 1, s_0 = 0) \\ &= (1 - p_s) \cdot p_0 + p_G \cdot (1 - p_0) \end{aligned}$$

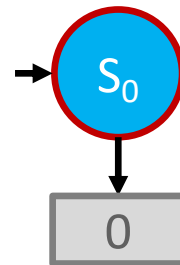
$$p(o_0 = 0) = p_s \cdot p_0 + (1 - p_G) \cdot (1 - p_0)$$



# Inference in BKT models

$$\underbrace{p(s_0 = 1|o_0 = 0)}_{p_{s_0|0}} = \frac{p(o_0 = 0|s_0 = 1) \cdot p(s_0 = 1)}{p(o_0 = 0)}$$
$$= \frac{p_s \cdot p_0}{p_s \cdot p_0 + (1 - p_G) \cdot (1 - p_0)}$$

$$p(s_0 = 0|o_0 = 0) = 1 - p_{s_0|0}$$

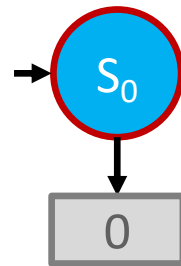




# Inference in BKT models

$$\underbrace{p(s_0 = 1 | o_0 = 1)}_{p_{s_0|1}} = \frac{p(o_0 = 1 | s_0 = 1) \cdot p(s_0 = 1)}{p(o_0 = 1)}$$
$$= \frac{(1 - p_s) \cdot p_0}{(1 - p_s) \cdot p_0 + p_G \cdot (1 - p_0)}$$

$$p(s_0 = 0 | o_0 = 1) = 1 - p_{s_0|1}$$

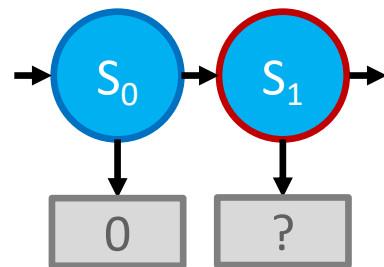


# Inference in BKT models

## Equations for time step 1:

$$\begin{aligned} p(s_1 = 1 | o_0 = 0) &= \frac{p(s_1 = 1, o_0 = 0)}{p(o_0 = 0)} \\ &= \frac{p(s_1 = 1, s_0 = 1, o_0 = 0)}{p(o_0 = 0)} + \frac{p(s_1 = 1, s_0 = 0, o_0 = 0)}{p(o_0 = 0)} \\ &= \frac{p(s_1 = 1 | s_0 = 1) \cdot p(o_0 = 0 | s_0 = 1) \cdot p(s_0 = 1)}{p(o_0 = 0)} \\ &\quad + \frac{p(s_1 = 1 | s_0 = 0) \cdot p(o_0 = 0 | s_0 = 0) \cdot p(s_0 = 0)}{p(o_0 = 0)} \end{aligned}$$

$$p(s_1 = 1 | o_0 = 0) = (1 - p_F) \cdot p_{s_0|0} + p_L \cdot (1 - p_{s_0|0})$$



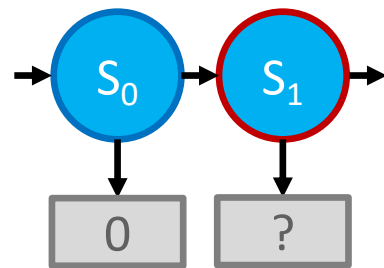
# Inference in BKT models

$$p(s_1 = 1|o_0 = 1) = (1 - p_F) \cdot p_{s_0|1} + p_L \cdot (1 - p_{s_0|1})$$

$$p(s_1 = 1|o_0 = 0) = (1 - p_F) \cdot p_{s_0|0} + p_L \cdot (1 - p_{s_0|0})$$



$$p_{s_1|o_0} = (1 - p_F) \cdot p_{s_0|o_0} + p_L \cdot (1 - p_{s_0|o_0})$$



# Inference in BKT models

$$p(o_1 = 1|o_0 = 0) = \frac{p(o_1 = 1, o_0 = 0)}{p(o_0 = 0)}$$

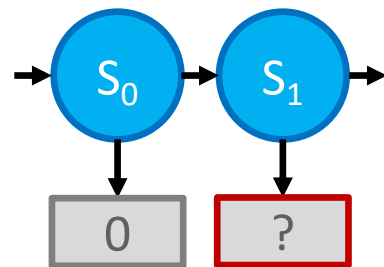
$$+ \frac{p(o_1 = 1, s_1 = 1, s_0 = 1, o_0 = 0)}{p(o_0 = 0)}$$

$$+ \frac{p(o_1 = 1, s_1 = 1, s_0 = 0, o_0 = 0)}{p(o_0 = 0)}$$

$$+ \frac{p(o_1 = 1, s_1 = 0, s_0 = 1, o_0 = 0)}{p(o_0 = 0)} + \frac{p(o_1 = 1, s_1 = 0, s_0 = 0, o_0 = 0)}{p(o_0 = 0)}$$

$$\begin{aligned} &= p(o_1 = 1|s_1 = 1) \cdot (p(s_1 = 1|s_0 = 1) \cdot p(o_0 = 0|s_0 = 1) \cdot p(s_0 = 1)/p(o_0 = 0)) \\ &+ p(o_1 = 1|s_1 = 1) \cdot (p(s_1 = 1|s_0 = 0) \cdot p(o_0 = 0|s_0 = 0) \cdot p(s_0 = 0)/p(o_0 = 0)) \\ &+ p(o_1 = 1|s_1 = 0) \cdot (p(s_1 = 0|s_0 = 1) \cdot p(o_0 = 0|s_0 = 1) \cdot p(s_0 = 1)/p(o_0 = 0)) \\ &+ p(o_1 = 1|s_1 = 0) \cdot (p(s_1 = 0|s_0 = 0) \cdot p(o_0 = 0|s_0 = 0) \cdot p(s_0 = 0)/p(o_0 = 0)) \end{aligned}$$

$$p(o_1 = 1|o_0 = 0) = (1 - p_S) \cdot p_{s_1|o_0=0} + p_G \cdot (1 - p_{s_1|o_0=0})$$



# Inference in BKT models

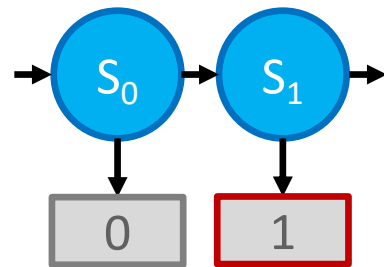
$$p(o_1 = 1|o_0 = 1) = (1 - p_S) \cdot p_{s_1|o_0=1} + p_G \cdot (1 - p_{s_1|o_0=1})$$

$$p(o_1 = 0|o_0 = 1) = p_S \cdot p_{s_1|o_0=1} + (1 - p_G) \cdot (1 - p_{s_1|o_0=1})$$



$$p(o_1 = 1|o_0) = (1 - p_S) \cdot p_{s_1|o_0} + p_G \cdot (1 - p_{s_1|o_0})$$

$$p(o_1 = 0|o_0) = p_S \cdot p_{s_1|o_0} + (1 - p_G) \cdot (1 - p_{s_1|o_0})$$



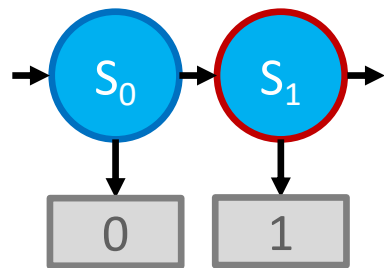
# Inference in BKT models

$$p(s_1 = 1 | o_1 = 1, o_0) = \frac{p(s_1 = 1, o_1 = 1, o_0)}{p(o_1 = 1, o_0)}$$

$$\begin{aligned} p(o_1 = 1, o_0) &= p(o_1 = 1 | o_0) \cdot p(o_0) \\ &= ((1 - p_S) \cdot p_{s_1|o_0} + p_G \cdot (1 - p_{s_1|o_0})) \cdot p(o_0) \end{aligned}$$

$$\begin{aligned} p(s_1 = 1, o_1 = 1, o_0) &= p(o_1 = 1 | s_1 = 1) \cdot p(s_1 = 1 | s_0 = 1) \cdot p(o_0 | s_0 = 1) \cdot p(s_0 = 1) \\ &\quad + p(o_1 = 1 | s_1 = 1) \cdot p(s_1 = 1 | s_0 = 0) \cdot p(o_0 | s_0 = 0) \cdot p(s_0 = 0) \\ &= (1 - p_S) \cdot p_{s_1|o_0} \cdot p(o_0) \end{aligned}$$

$$p(s_1 = 1 | o_1 = 1, o_0) = \frac{(1 - p_S) \cdot p_{s_1|o_0}}{((1 - p_S) \cdot p_{s_1|o_0} + p_G \cdot (1 - p_{s_1|o_0}))}$$



# Inference in BKT models

$$\mathbf{o}_{t-1} = [o_0, \dots, o_{t-1}]$$

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

$$p_{s_0|1} = \frac{(1 - p_S) \cdot p_0}{(1 - p_S) \cdot p_0 + p_G \cdot (1 - p_0)} \quad \left. \vphantom{p_{s_0|1}} \right\} p_{s_0|o_0}$$

$$p_{s_0|0} = \frac{p_S \cdot p_0}{p_S \cdot p_0 + (1 - p_G) \cdot (1 - p_0)} \quad \left. \vphantom{p_{s_0|0}} \right\} p_{s_0|o_0}$$

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

$$p_{s_t|\mathbf{o}_{t-1}} = (1 - p_F) \cdot p_{s_{t-1}|\mathbf{o}_{t-1}} + p_L \cdot (1 - p_{s_{t-1}|\mathbf{o}_{t-1}})$$

$$p(o_t = 1|\mathbf{o}_{t-1}) = (1 - p_S) \cdot p_{s_t|\mathbf{o}_{t-1}} + p_G \cdot (1 - p_{s_t|\mathbf{o}_{t-1}})$$

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

$$p_{s_t|1,\mathbf{o}_{t-1}} = \frac{(1 - p_S) \cdot p_{s_t|\mathbf{o}_{t-1}}}{(1 - p_S) \cdot p_{s_t|\mathbf{o}_{t-1}} + p_G \cdot (1 - p_{s_t|\mathbf{o}_{t-1}})} \quad \left. \vphantom{p_{s_t|1,\mathbf{o}_{t-1}}} \right\} p_{s_t|\mathbf{o}_t}$$

$$p_{s_t|0,\mathbf{o}_{t-1}} = \frac{p_S \cdot p_{s_t|\mathbf{o}_{t-1}}}{p_S \cdot p_{s_t|\mathbf{o}_{t-1}} + (1 - p_G) \cdot (1 - p_{s_t|\mathbf{o}_{t-1}})} \quad \left. \vphantom{p_{s_t|0,\mathbf{o}_{t-1}}} \right\} p_{s_t|\mathbf{o}_t}$$

# Making predictions using a BKT model

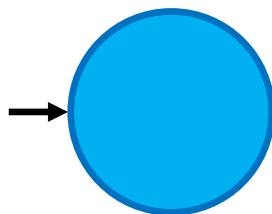
$$p_0 = 0.5$$

$$p_S = 0.2$$

$$p_G = 0.3$$

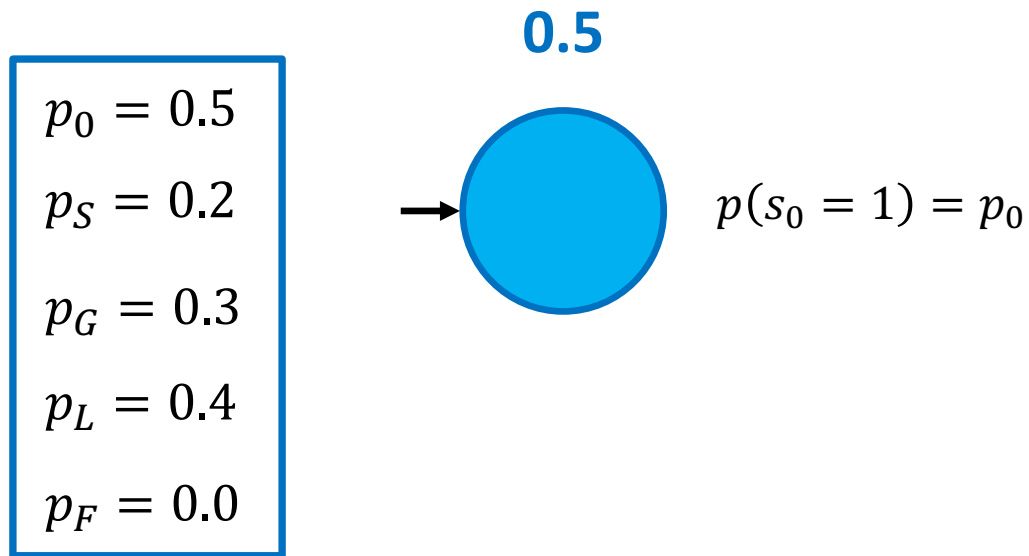
$$p_L = 0.4$$

$$p_F = 0.0$$

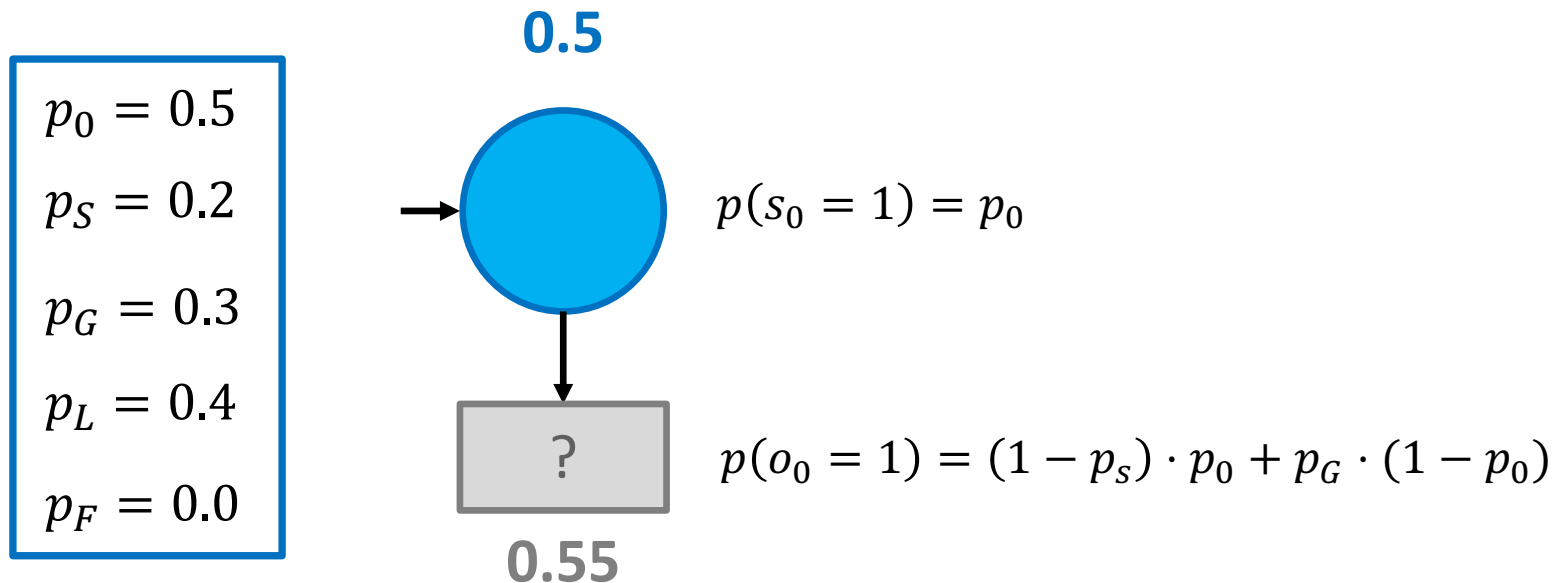




# Making predictions using a BKT model

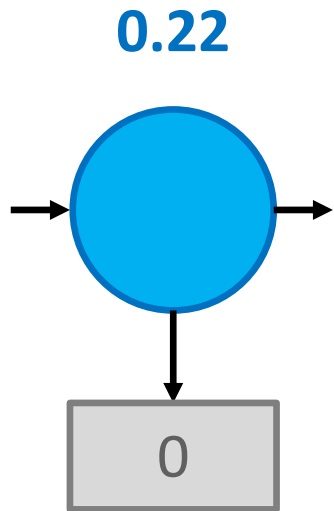


# Making predictions using a BKT model



# Making predictions using a BKT model

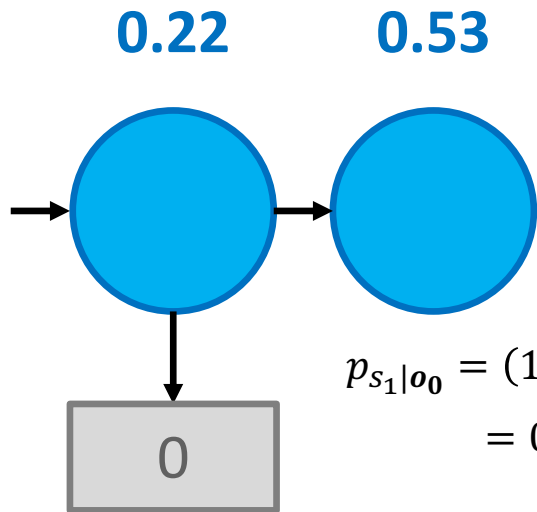
$$\begin{aligned}p_0 &= 0.5 \\p_S &= 0.2 \\p_G &= 0.3 \\p_L &= 0.4 \\p_F &= 0.0\end{aligned}$$



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

# Making predictions using a BKT model

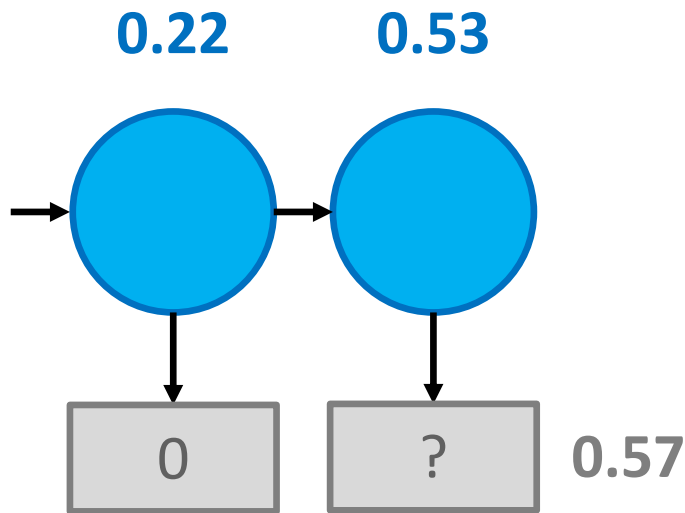
$$\begin{aligned}p_0 &= 0.5 \\p_S &= 0.2 \\p_G &= 0.3 \\p_L &= 0.4 \\p_F &= 0.0\end{aligned}$$



$$\begin{aligned}p_{s_1|o_0} &= (1 - p_F) \cdot p_{s_1|o_0} + p_L \cdot (1 - p_{s_1|o_0}) \\&= 0.22 + 0.4 \cdot 0.78\end{aligned}$$

# Making predictions using a BKT model

$$\begin{aligned} p_0 &= 0.5 \\ p_S &= 0.2 \\ p_G &= 0.3 \\ p_L &= 0.4 \\ p_F &= 0.0 \end{aligned}$$



$$\begin{aligned} p(o_1 = 1 | \mathbf{o}_0) &= (1 - p_S) \cdot p_{s_1 | \mathbf{o}_0} + p_G \cdot (1 - p_{s_1 | \mathbf{o}_0}) \\ &= 0.8 \cdot 0.53 + 0.3 \cdot 0.47 \end{aligned}$$

# Making predictions using a BKT model

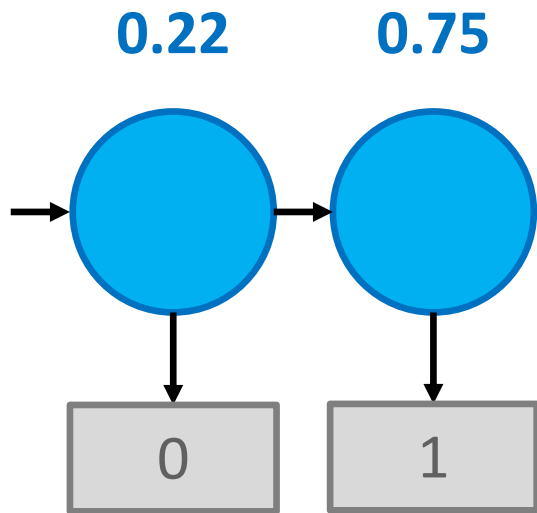
$$p_0 = 0.5$$

$$p_S = 0.2$$

$$p_G = 0.3$$

$$p_L = 0.4$$

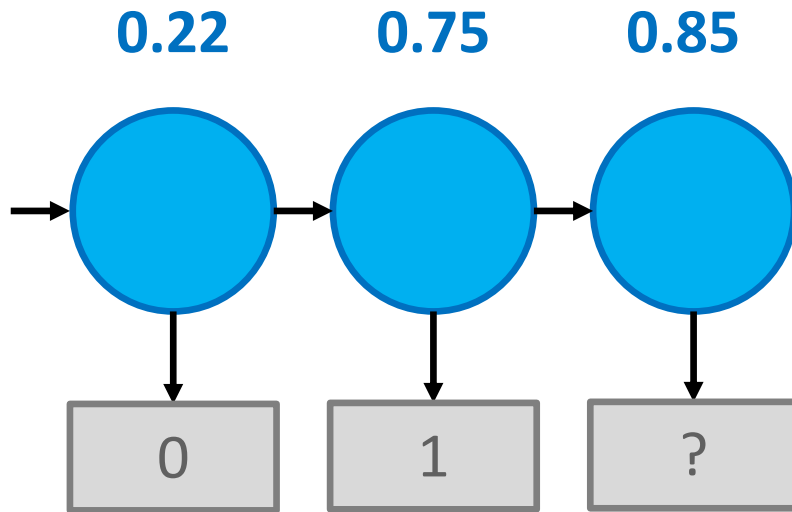
$$p_F = 0.0$$



$$p_{s_1|1,o_0} = \frac{(1 - p_S) \cdot p_{s_1|o_0}}{(1 - p_S) \cdot p_{s_1|o_0} + p_G \cdot (1 - p_{s_1|o_0})}$$
$$= \frac{0.8 \cdot 53}{0.57}$$

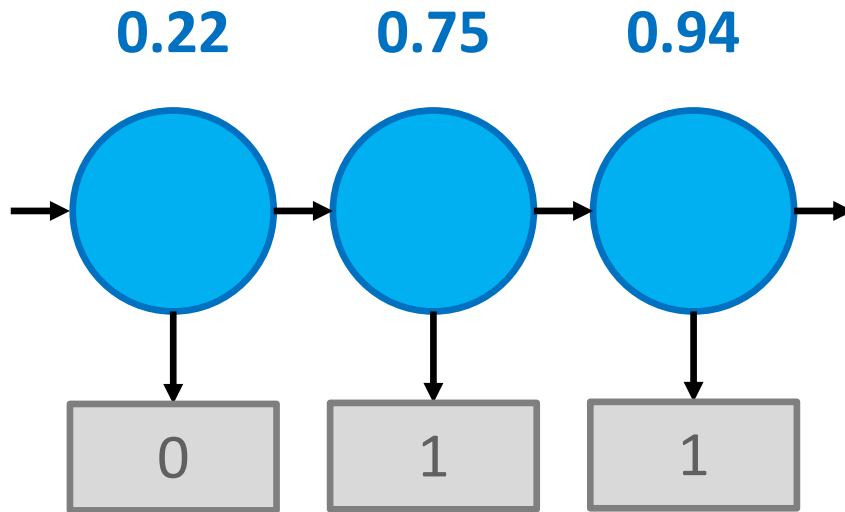
# Making predictions using a BKT model

$$\begin{aligned} p_0 &= 0.5 \\ p_S &= 0.2 \\ p_G &= 0.3 \\ p_L &= 0.4 \\ p_F &= 0.0 \end{aligned}$$



# Making predictions using a BKT model

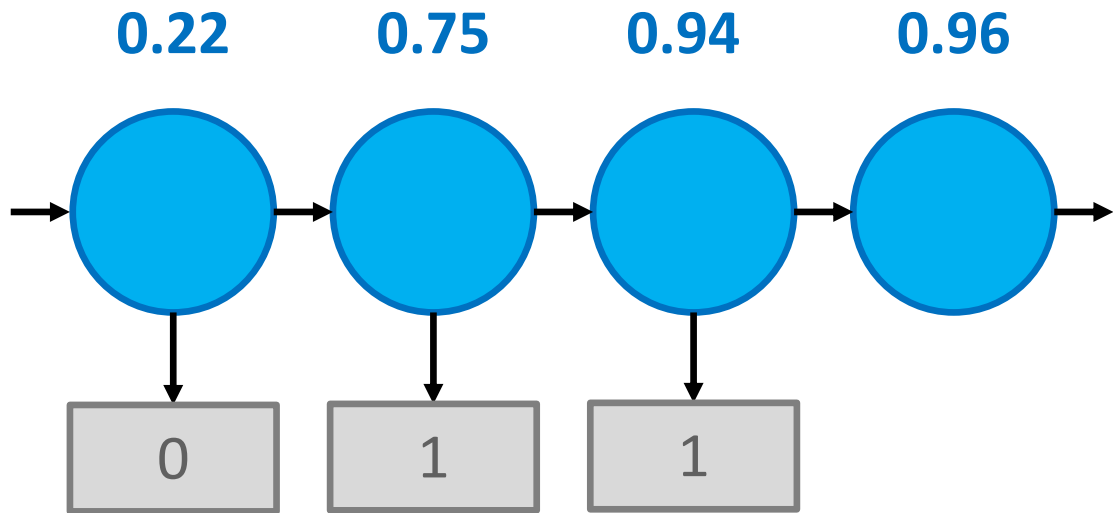
$$\begin{aligned} p_0 &= 0.5 \\ p_S &= 0.2 \\ p_G &= 0.3 \\ p_L &= 0.4 \\ p_F &= 0.0 \end{aligned}$$





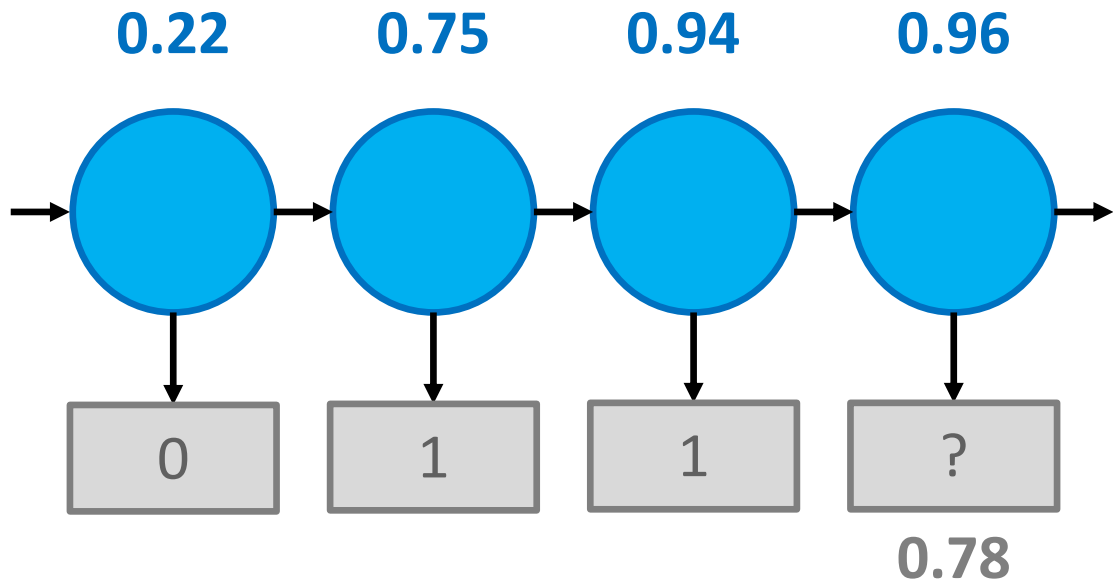
# Making predictions using a BKT model

$$\begin{aligned} p_0 &= 0.5 \\ p_S &= 0.2 \\ p_G &= 0.3 \\ p_L &= 0.4 \\ p_F &= 0.0 \end{aligned}$$



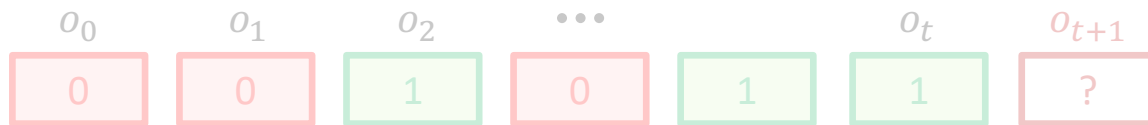
# Making predictions using a BKT model

$$\begin{aligned} p_0 &= 0.5 \\ p_S &= 0.2 \\ p_G &= 0.3 \\ p_L &= 0.4 \\ p_F &= 0.0 \end{aligned}$$



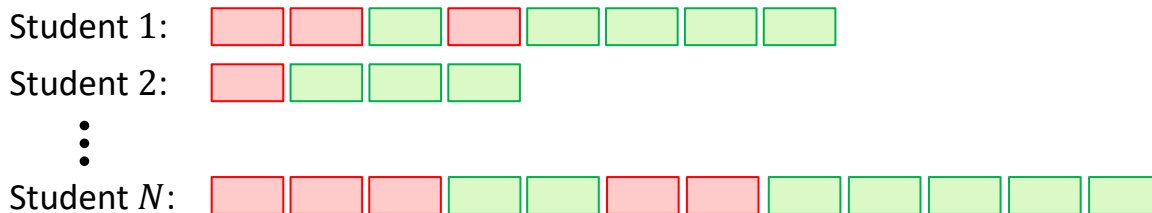
# Two tasks need to be solved in practice

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



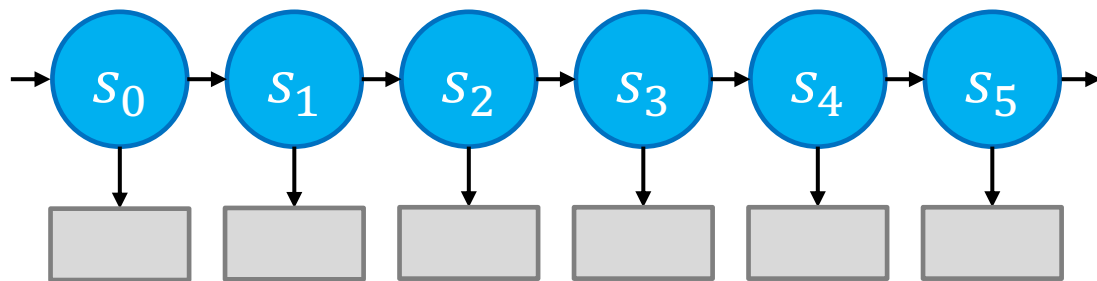
Inference

- Given sequences of observations  $\mathbf{o} = [o_0, \dots, o_T]$  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

# Training a BKT model



$$\theta = \{p_0, p_L, p_F, p_S, p_G\}$$

Student  $l_0$ :  $\mathbf{o}_{l_0} = [0, 1, 1]$

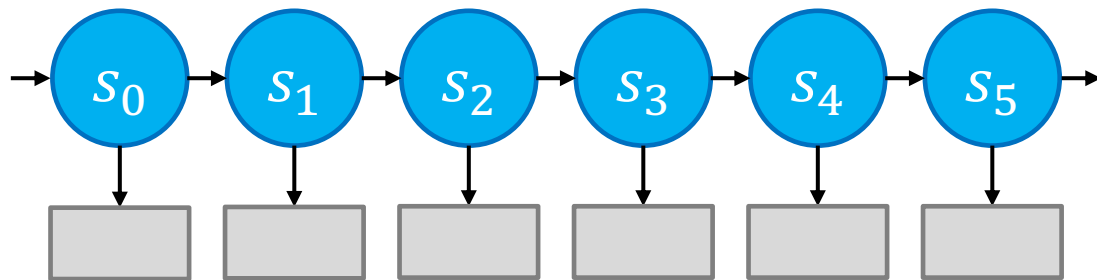
$\vdots$

Student  $l_{N-1}$ :  $\mathbf{o}_{l_{N-1}} = [1, 0, 1, 1, 1, 0, 0, 1, 1, 1]$

Student  $l_N$ :  $\mathbf{o}_{l_N} = [0, 1, 0, 1]$

$$\max_{\theta} p(\mathbf{o}_{l_0}, \dots, \mathbf{o}_{l_{N-1}}, \mathbf{o}_{l_N})$$

# Training a BKT model



$$\theta = \{p_0, p_L, p_F, p_S, p_G\}$$

Student  $l_0$ :  $\mathbf{o}_{l_0} = [0, 1, 1]$

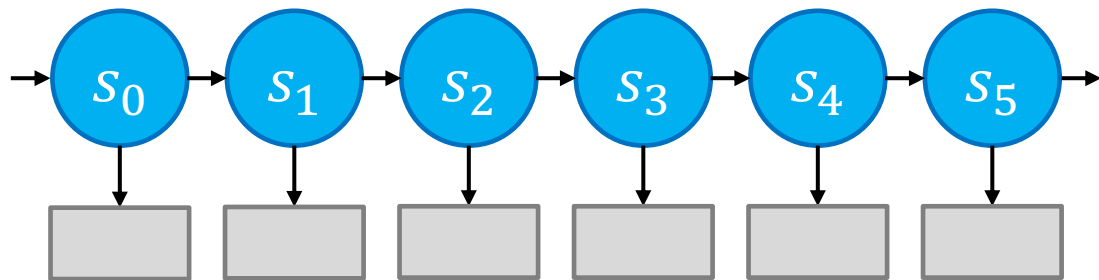
$\vdots$

Student  $l_{N-1}$ :  $\mathbf{o}_{l_{N-1}} = [1, 0, 1, 1, 1, 0, 0, 1, 1, 1]$

Student  $l_N$ :  $\mathbf{o}_{l_N} = [0, 1, 0, 1]$

$$\max_{\theta} \prod_{i=1}^N p(\mathbf{o}_{l_i})$$

# Training a BKT model



$$\theta = \{p_0, p_L, p_F, p_S, p_G\}$$

$$\text{Student } l_0: \quad p(\mathbf{o}_{l_0}) = \sum_s p(\mathbf{o}_{l_0}, s)$$

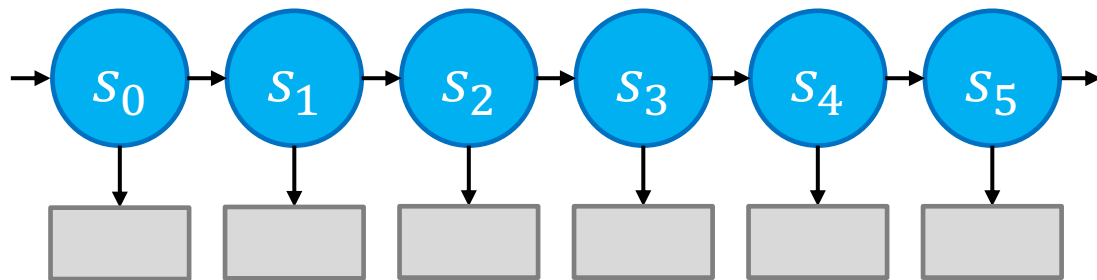
$\vdots$

$$\text{Student } l_{N-1}: \quad p(\mathbf{o}_{l_0}) = \sum_s p(\mathbf{o}_{l_0}, s)$$

$$\text{Student } l_N: \quad p(\mathbf{o}_{l_0}) = \sum_s p(\mathbf{o}_{l_0}, s)$$

$$\max_{\theta} \prod_{i=1}^N \sum_{s_{l_i}} p(\mathbf{o}_{l_i}, s_{l_i})$$

# Training a BKT model



$$\theta = \{p_0, p_L, p_F, p_S, p_G\}$$

$$\max_{\theta} \prod_{i=1}^N \sum_{s_{l_i}} p(\mathbf{o}_{l_i}, s_{l_i}) \quad \rightarrow \quad \min_{\theta} - \sum_{i=1}^N \log \left( \sum_{s_{l_i}} p(\mathbf{o}_{l_i}, s_{l_i}) \right)$$

- Brute-Force Grid Search
- Expectation Maximization
- Gradient Descent
- Nelder-Mead Optimization

# Your Turn – Evaluating a BKT model

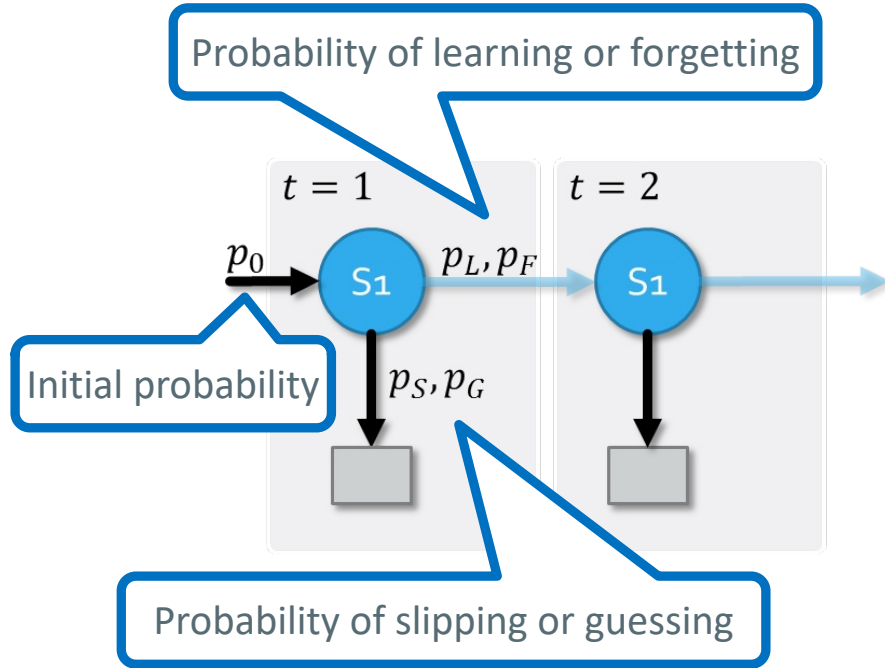
- In the student notebook, you have:
    - A trained BKT model for six selected skills
    - A data frame containing the predictions of the BKT model for each observation in the test set
  - Your task:
    - Compute the RMSE or AUC separately for each skill
    - Provide a visualization of the mean RMSE (or AUC) + standard deviation over all skills as well as the per skill RMSE (or AUC)
-



# Summary – Why tracing 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)
  - ➡ Choose the next appropriate activity
  - ➡ Know which activities support learning
-

# Summary - BKT



- Predict  $p(o_{i_{s_1},t} | o_0, \dots, o_{t-1})$ , the probability that the student will solve task  $i_{s_1}$  correctly at time step  $t$
- Predict  $p(s_{1,t} | o_0, \dots, o_{t-1})$ , the probability that the student has mastered skill  $s_1$  at time step  $t$

# Summary - Assumptions behind BKT

- Knowledge can be divided into different skills
  - Definition of skills is accurate/detailed enough
  - Each task corresponds to a single skill (original)
  - There is **no** connection between the skills
  - Mastery can be achieved through practice
  - There is no forgetting:  $p_F = 0$  (original)
-