Computerized adaptive testing using Bayesian networks

Jiří Vomlel

Academy of Sciences of the Czech Republic

11th July, 2007

Educational Testing Service (ETS)

- ETS is the world's largest private educational testing organization. It has 2300 permanent employees.

Educational Testing Service (ETS)

- ETS is the world's largest private educational testing organization. It has 2300 permanent employees.
- Number of participants in tests in the school year 2000/2001:

3 185 000	SAT I Reasoning Test and SAT II: Subject Area Tests
2 293 000	PSAT: Preliminary SAT/National Merit Scholarship Qualifying Test
1 421 000	AP: Advanced Placement Program
801 000	The Praxis Series: Professional Assessments for Beginning Teachers and Pre-Professional Skills Tests
787 000	TOEFL: Test of English as a Foreign Language
449 000	GRE: Graduate Record Examinations General Test

Educational Testing Service (ETS)

- ETS is the world's largest private educational testing organization. It has 2300 permanent employees.
- Number of participants in tests in the school year 2000/2001:

3 185 000	SAT I Reasoning Test and SAT II: Subject Area Tests
2 293 000	PSAT: Preliminary SAT/National Merit Scholarship Qualifying Test
1 421 000	AP: Advanced Placement Program
801 000	The Praxis Series: Professional Assessments for Beginning Teachers and Pre-Professional Skills Tests
787 000	TOEFL: Test of English as a Foreign Language
449 000	GRE: Graduate Record Examinations General Test

- ETS has a research unit doing research on adaptive testing using Bayesian networks: `http://www.ets.org/research/`

Rasch model (G. Rasch,1960)

Model variables

$Y_{n,i}$	binary response variable - its values indicates whether the answer of person n to question i was correct
$n = 1, \ldots, N$	person index
$i = 1, \ldots, I$	question index

Rasch model (G. Rasch,1960)

Model variables

$Y_{n,i}$	binary response variable - its values indicates whether the answer of person n to question i was correct
$n = 1, \ldots, N$	person index
$i = 1, \ldots, I$	question index

Model parameters

δ_i	difficulty of question i - fixed effects
β_n	ability (knowledge level) of person n - a random effect

Models for the response variable Y

$$Y_{n,i} = \begin{cases} 1 & \text{if } \beta_n \geq \delta_i \\ 0 & \text{otherwise.} \end{cases}$$

Models for the response variable Y

$$Y_{n,i} = \begin{cases} 1 & \text{if } \beta_n \geq \delta_i \\ 0 & \text{otherwise.} \end{cases}$$

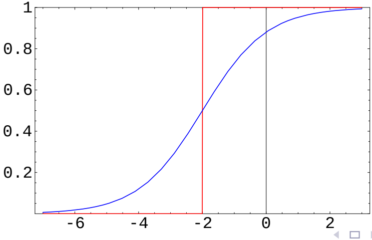
$$P(Y_{n,i} = 1) = \frac{\exp(\beta_n - \delta_i)}{1 + \exp(\beta_n - \delta_i)}$$

Models for the response variable Y

$$Y_{n,i} = \begin{cases} 1 & \text{if } \beta_n \geq \delta_i \\ 0 & \text{otherwise.} \end{cases}$$

$$P(Y_{n,i} = 1) = \frac{\exp(\beta_n - \delta_i)}{1 + \exp(\beta_n - \delta_i)}$$

$P(Y_{n,i} = 1 \mid \beta_n)$ for $\delta_i = -2$



Probability distribution for random effect β_n

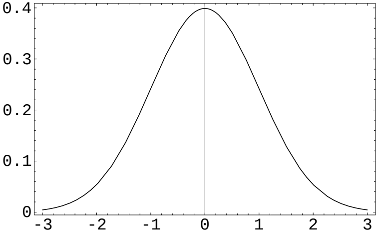
$$P(\beta_n) \quad = \quad \mathcal{N}(0, \sigma^2)$$

a normal (Gaussian) distribution
with the mean equal zero, and variance σ^2.

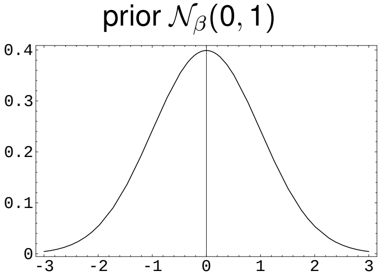
Probability distribution for random effect β_n

$$P(\beta_n) \quad = \quad \mathcal{N}(0, \sigma^2)$$

a normal (Gaussian) distribution
with the mean equal zero, and variance σ^2.



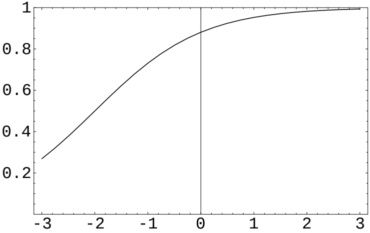
Computations with the Rasch model



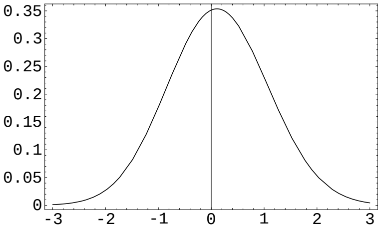
$\mathcal{N}_\beta(0,1)$

Computations with the Rasch model

$P(Y = 1 \mid \beta, \delta_1 = -2)$



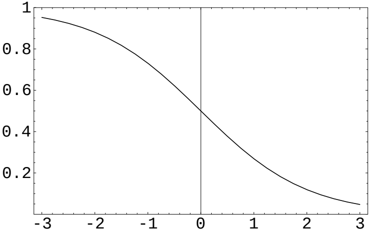
posterior



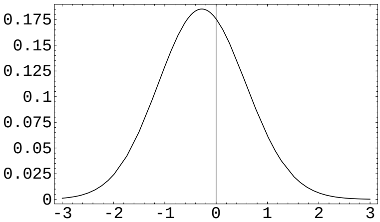
$$\mathcal{N}_\beta(0, 1) \cdot P(Y = 1 \mid \beta, \delta_1 = -2)$$

Computations with the Rasch model

$P(Y = 0 \mid \beta, \delta_2 = 0)$

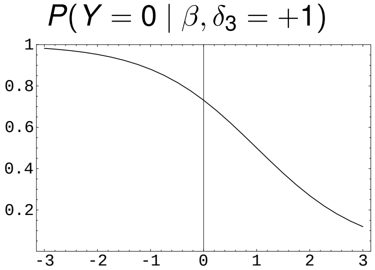


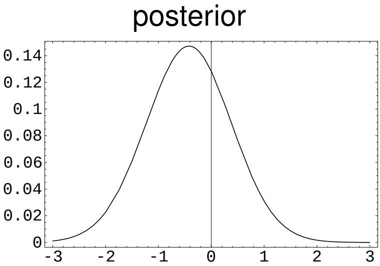
posterior



$$\mathcal{N}_\beta(0,1) \cdot P(Y = 1 \mid \beta, \delta_1 = -2) \cdot P(Y = 0 \mid \beta, \delta_2 = 0)$$

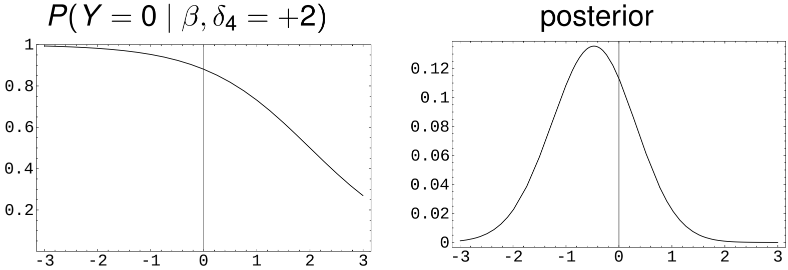
Computations with the Rasch model





$$\mathcal{N}_\beta(0,1) \cdot P(Y = 1 \mid \beta, \delta_1 = -2) \cdot P(Y = 0 \mid \beta, \delta_2 = 0)$$
$$\cdot P(Y = 0 \mid \beta, \delta_3 = +1)$$

Computations with the Rasch model



$$\mathcal{N}_\beta(0,1) \cdot P(Y = 1 \mid \beta, \delta_1 = -2) \cdot P(Y = 0 \mid \beta, \delta_2 = 0)$$
$$\cdot P(Y = 0 \mid \beta, \delta_3 = +1) \cdot P(Y = 0 \mid \beta, \delta_4 = +2)$$

Student model and evidence models (R. Almond and R. Mislevy, 1999)

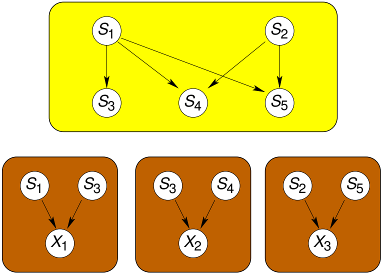
- The variables of the models are: (1) skills, abilities, misconceptions, etc. - for brevity called skills - the vector of skills is denoted $\boldsymbol{S}$ and (2) items (questions) - the vector of questions is denoted $\boldsymbol{X}$.

Student model and evidence models (R. Almond and R. Mislevy, 1999)

- The variables of the models are: (1) skills, abilities, misconceptions, etc. - for brevity called skills - the vector of skills is denoted $\boldsymbol{S}$ and (2) items (questions) - the vector of questions is denoted $\boldsymbol{X}$.
- **student model** describes relations between student skills.

Student model and evidence models (R. Almond and R. Mislevy, 1999)

- The variables of the models are: (1) skills, abilities, misconceptions, etc. - for brevity called skills - the vector of skills is denoted $\boldsymbol{S}$ and (2) items (questions) - the vector of questions is denoted $\boldsymbol{X}$.
- **student model** describes relations between student skills.
- **evidence models** - one for each item (question) - describes relations of the item to the skills.



Building Bayesian network models

Three basic approaches:

Building Bayesian network models

Three basic approaches:

- Discussions with **domain experts**: expert knowledge is used to get the structure and parameters of the model

Building Bayesian network models

Three basic approaches:

- Discussions with **domain experts**: expert knowledge is used to get the structure and parameters of the model
- A dataset of records is collected and a **machine learning** method is used to to construct a model and estimate its parameters.

Building Bayesian network models

Three basic approaches:

- Discussions with **domain experts**: expert knowledge is used to get the structure and parameters of the model
- A dataset of records is collected and a **machine learning** method is used to to construct a model and estimate its parameters.
- A **combination** of previous two: e.g. experts suggest the structure and collected data are used to estimate parameters.

Example of a simple diagnostic task

We want to diagnose presence/absence of three skills

$$S_1, S_2, S_3$$

Example of a simple diagnostic task

We want to diagnose presence/absence of three skills

$$S_1, S_2, S_3$$

using three questions

$$X_{1,2}, X_{1,3}, X_{2,3} \quad .$$

Example of a simple diagnostic task

We want to diagnose presence/absence of three skills

$$S_1, S_2, S_3$$

using three questions

$$X_{1,2}, X_{1,3}, X_{2,3} \quad .$$

The questions depend on skills and their dependence is described by conditional probability distributions

$$P(X_{i,j} = 1 | S_i = s_i, S_j = s_j) \quad = \quad \begin{cases} 1 & \text{if } (s_i, s_j) = (1, 1) \\ 0 & \text{otherwise.} \end{cases}$$

Example of a simple diagnostic task

We want to diagnose presence/absence of three skills

$$S_1, S_2, S_3$$

using three questions

$$X_{1,2}, X_{1,3}, X_{2,3} \quad .$$

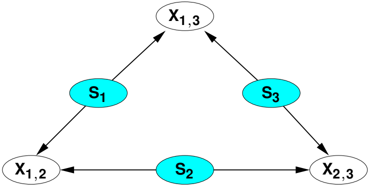
The questions depend on skills and their dependence is described by conditional probability distributions

$$P(X_{i,j} = 1 | S_i = s_i, S_j = s_j) \quad = \quad \begin{cases} 1 & \text{if } (s_i, s_j) = (1,1) \\ 0 & \text{otherwise.} \end{cases}$$

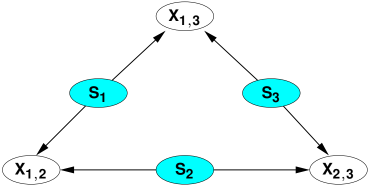
Assume all answers were wrong, i.e.,

$$X_{1,2} = 0, \quad X_{1,3} = 0, \quad X_{2,3} = 0 \quad .$$

Reasoning under the assumption of skills' independence



Reasoning under the assumption of skills' independence

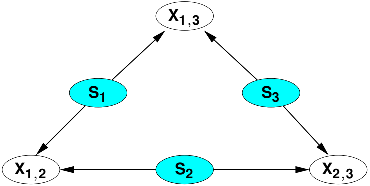


First, assume the skills are pairwise independent, i.e.,

$$P(S_1, S_2, S_3) = P(S_1) \cdot P(S_2) \cdot P(S_3)$$

and for $i = 1, 2, 3$, $s_i = 0, 1$
$P(S_i = s_i) = 0.5$

Reasoning under the assumption of skills' independence



First, assume the skills are pairwise independent, i.e.,

$$P(S_1, S_2, S_3) = P(S_1) \cdot P(S_2) \cdot P(S_3)$$

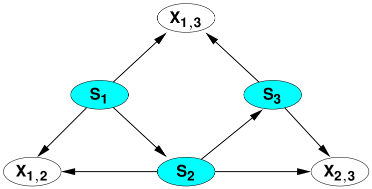
and for $i = 1, 2, 3$, $s_i = 0, 1$
$P(S_i = s_i) = 0.5$

Then conditional probabilities for $j = 1, 2, 3$ are

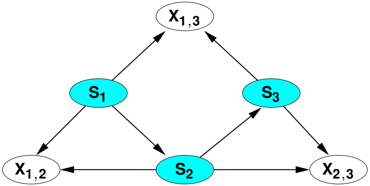
$$P(S_j = 0 \mid X_{1,2} = 0, X_{1,3} = 0, X_{2,3} = 0) = 0.75 \quad ,$$

i.e., we cannot decide with certainty, which skills are present and which are absent.

Modeling dependence between skills



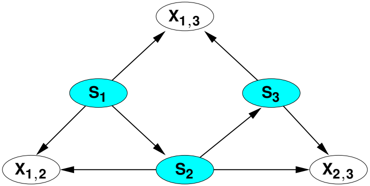
Modeling dependence between skills



Now, assume there is a deterministic hierarchy among skills

$$S_1 \Rightarrow S_2, \; S_2 \Rightarrow S_3$$

Modeling dependence between skills



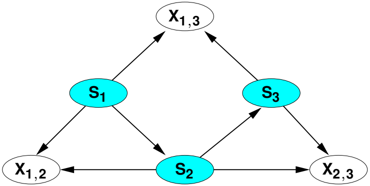
Now, assume there is a deterministic hierarchy among skills

$$S_1 \Rightarrow S_2, \; S_2 \Rightarrow S_3$$

Then conditional probabilities are

$$\begin{aligned}
P(S_1 = 0 \mid X_{1,2} = 0, X_{1,3} = 0, X_{2,3} = 0) &= 1 \\
P(S_2 = 0 \mid X_{1,2} = 0, X_{1,3} = 0, X_{2,3} = 0) &= 1 \\
P(S_3 = 0 \mid X_{1,2} = 0, X_{1,3} = 0, X_{2,3} = 0) &= 0.5
\end{aligned}$$

Modeling dependence between skills



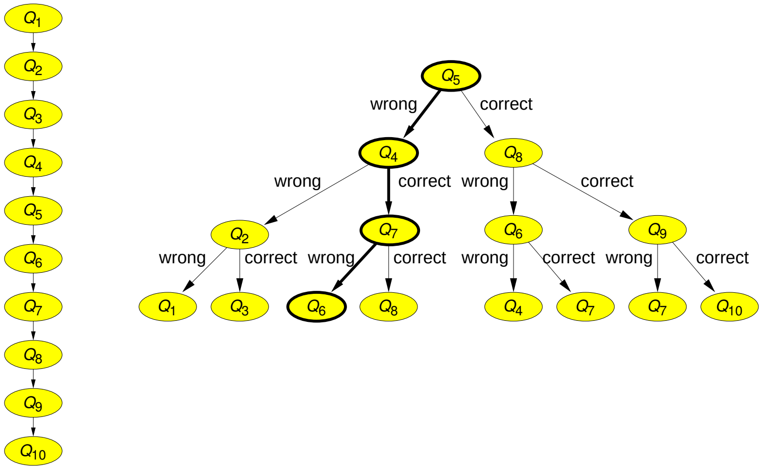
Now, assume there is a deterministic hierarchy among skills

$$S_1 \Rightarrow S_2, \; S_2 \Rightarrow S_3$$

Then conditional probabilities are

$$\begin{aligned} P(S_1 = 0 \mid X_{1,2} = 0, X_{1,3} = 0, X_{2,3} = 0) &= 1 \\ P(S_2 = 0 \mid X_{1,2} = 0, X_{1,3} = 0, X_{2,3} = 0) &= 1 \\ P(S_3 = 0 \mid X_{1,2} = 0, X_{1,3} = 0, X_{2,3} = 0) &= 0.5 \end{aligned}$$

For $i = 1, 2, 3$: $P(S_i \mid X_{1,2} = 0, X_{1,3} = 0, X_{2,3} = 0) = P(S_i \mid X_{2,3} = 0)$.
$X_{2,3} = 0$ gives **the same information** as $X_{1,2} = X_{1,3} = X_{2,3} = 0$.


Q_1
Q_2
Q_3
Q_4
Q_5
Q_6
Q_7
Q_8
Q_9
Q_{10}
Q_5
wrong
correct
Q_4
Q_8
wrong
correct
wrong
correct
Q_2
Q_7
Q_6
Q_9
wrong
correct
wrong
correct
wrong
correct
wrong
correct
Q_1
Q_3
Q_6
Q_8
Q_4
Q_7
Q_7
Q_{10}

Computerized Adaptive Testing (CAT)

The goal of CAT is to tailor each test so that it brings most information about each student.

Computerized Adaptive Testing (CAT)

The goal of CAT is to tailor each test so that it brings most information about each student.

Two basic steps are repeated:

1. estimation of the knowledge level of the tested student

Computerized Adaptive Testing (CAT)

The goal of CAT is to tailor each test so that it brings most information about each student.

Two basic steps are repeated:

1. estimation of the knowledge level of the tested student
2. selection of appropriate question to ask the student

Computerized Adaptive Testing (CAT)

The goal of CAT is to tailor each test so that it brings most information about each student.

Two basic steps are repeated:

1. estimation of the knowledge level of the tested student
2. selection of appropriate question to ask the student

Entropy as an information measure:

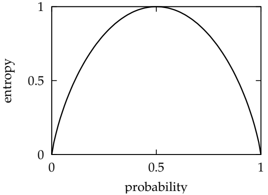
Computerized Adaptive Testing (CAT)

The goal of CAT is to tailor each test so that it brings most information about each student.
Two basic steps are repeated:

1. estimation of the knowledge level of the tested student
2. selection of appropriate question to ask the student

Entropy as an information measure:

- $H(P(\boldsymbol{S})) = -\sum_{\boldsymbol{s}} P(\boldsymbol{S}=\boldsymbol{s}) \cdot \log P(\boldsymbol{S}=\boldsymbol{s})$



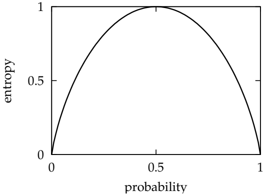
Computerized Adaptive Testing (CAT)

The goal of CAT is to tailor each test so that it brings most information about each student.
Two basic steps are repeated:

1. estimation of the knowledge level of the tested student
2. selection of appropriate question to ask the student

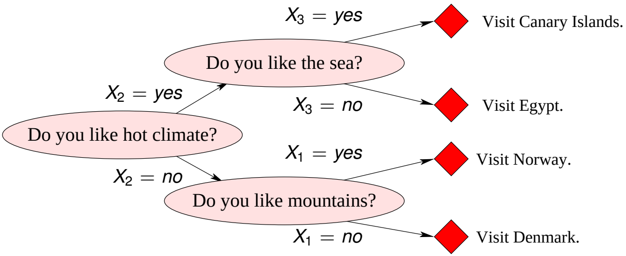
Entropy as an information measure:

- $H(P(\boldsymbol{S})) = -\sum_{\boldsymbol{s}} P(\boldsymbol{S} = \boldsymbol{s}) \cdot \log P(\boldsymbol{S} = \boldsymbol{s})$

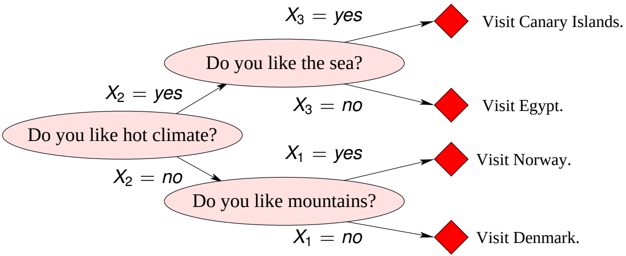


- "The lower the entropy the more we know."

Building strategies using the models



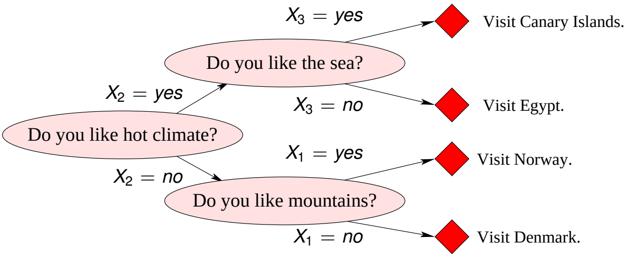
Building strategies using the models



For all terminal nodes (leaves) $\ell \in \mathcal{L}(\boldsymbol{s})$ of a strategy $\boldsymbol{s}$ we define:

- steps that were performed to get to that node (e.g. questions answered in a certain way). It is called collected evidence $\boldsymbol{e}_\ell$.

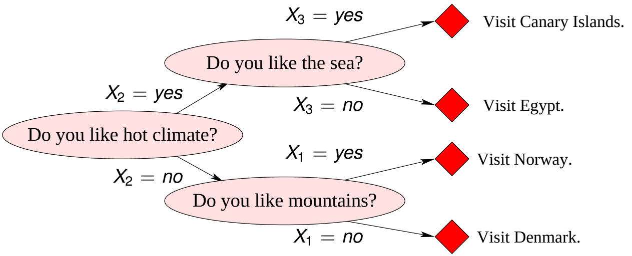
Building strategies using the models



For all terminal nodes (leaves) $\ell \in \mathcal{L}(\boldsymbol{s})$ of a strategy $\boldsymbol{s}$ we define:

- steps that were performed to get to that node (e.g. questions answered in a certain way). It is called collected evidence $\boldsymbol{e}_\ell$.
- Using the probabilistic model of the domain we can compute probability of getting to a terminal node $P(\boldsymbol{e}_\ell)$.

Building strategies using the models



For all terminal nodes (leaves) $\ell \in \mathcal{L}(\boldsymbol{s})$ of a strategy $\boldsymbol{s}$ we define:

- steps that were performed to get to that node (e.g. questions answered in a certain way). It is called collected evidence $\boldsymbol{e}_\ell$.
- Using the probabilistic model of the domain we can compute probability of getting to a terminal node $P(\boldsymbol{e}_\ell)$.
- Also during the process, when we collected evidence $\boldsymbol{e}$, we update the probability of getting to a terminal node to conditional probability $P(\boldsymbol{e}_\ell | \boldsymbol{e})$

Building strategies using the models

For all terminal nodes $\ell \in \mathcal{L}(\boldsymbol{s})$ of a strategy $\boldsymbol{s}$ we also define:

- an evaluation function $f : \cup_{\boldsymbol{s} \in \mathcal{S}} \mathcal{L}(\boldsymbol{s}) \mapsto \mathbb{R}$,

Building strategies using the models

For all terminal nodes $\ell \in \mathcal{L}(\boldsymbol{s})$ of a strategy $\boldsymbol{s}$ we also define:

- an evaluation function $f : \cup_{\boldsymbol{s} \in \mathcal{S}} \mathcal{L}(\boldsymbol{s}) \mapsto \mathbb{R}$,
- The evaluation function can be, e.g., the information gain.

Building strategies using the models

For all terminal nodes $\ell \in \mathcal{L}(\boldsymbol{s})$ of a strategy $\boldsymbol{s}$ we also define:

- an evaluation function $f : \cup_{\boldsymbol{s} \in \mathcal{S}} \mathcal{L}(\boldsymbol{s}) \mapsto \mathbb{R}$,
- The evaluation function can be, e.g., the information gain.
- Information gain in a node n of a strategy
  $IG(\boldsymbol{e}_n) = H(P(\boldsymbol{S})) - H(P(\boldsymbol{S} \mid \boldsymbol{e}_n))$

Building strategies using the models

For all terminal nodes $\ell \in \mathcal{L}(\boldsymbol{s})$ of a strategy $\boldsymbol{s}$ we also define:

- an evaluation function $f : \cup_{\boldsymbol{s} \in \mathcal{S}} \mathcal{L}(\boldsymbol{s}) \mapsto \mathbb{R}$,
- The evaluation function can be, e.g., the information gain.
- Information gain in a node n of a strategy
  $IG(\boldsymbol{e}_n) = H(P(\boldsymbol{S})) - H(P(\boldsymbol{S} \mid \boldsymbol{e}_n))$

For each strategy we can compute:

Building strategies using the models

For all terminal nodes $\ell \in \mathcal{L}(\boldsymbol{s})$ of a strategy $\boldsymbol{s}$ we also define:

- an evaluation function $f : \cup_{\boldsymbol{s} \in \mathcal{S}} \mathcal{L}(\boldsymbol{s}) \mapsto \mathbb{R}$,
- The evaluation function can be, e.g., the information gain.
- Information gain in a node n of a strategy
  $IG(\boldsymbol{e}_n) = H(P(\boldsymbol{S})) - H(P(\boldsymbol{S} \mid \boldsymbol{e}_n))$

For each strategy we can compute:

- expected value of the strategy:

$$E_f(\boldsymbol{s}) = \sum_{\ell \in \mathcal{L}(\boldsymbol{s})} P(\boldsymbol{e}_\ell) \cdot f(\boldsymbol{e}_\ell)$$

Building strategies using the models

For all terminal nodes $\ell \in \mathcal{L}(\boldsymbol{s})$ of a strategy $\boldsymbol{s}$ we also define:

- an evaluation function $f : \cup_{\boldsymbol{s} \in \mathcal{S}} \mathcal{L}(\boldsymbol{s}) \mapsto \mathbb{R}$,
- The evaluation function can be, e.g., the information gain.
- Information gain in a node n of a strategy $IG(\boldsymbol{e}_n) = H(P(\boldsymbol{S})) - H(P(\boldsymbol{S} \mid \boldsymbol{e}_n))$

For each strategy we can compute:

- expected value of the strategy:

$$E_f(\boldsymbol{s}) \quad = \quad \sum_{\ell \in \mathcal{L}(\boldsymbol{s})} P(\boldsymbol{e}_\ell) \cdot f(\boldsymbol{e}_\ell)$$

The goal is

Building strategies using the models

For all terminal nodes $\ell \in \mathcal{L}(\boldsymbol{s})$ of a strategy $\boldsymbol{s}$ we also define:

- an evaluation function $f : \cup_{\boldsymbol{s} \in \mathcal{S}} \mathcal{L}(\boldsymbol{s}) \mapsto \mathbb{R}$,
- The evaluation function can be, e.g., the information gain.
- Information gain in a node n of a strategy $IG(\boldsymbol{e}_n) = H(P(\boldsymbol{S})) - H(P(\boldsymbol{S} \mid \boldsymbol{e}_n))$

For each strategy we can compute:

- expected value of the strategy:

$$E_f(\boldsymbol{s}) = \sum_{\ell \in \mathcal{L}(\boldsymbol{s})} P(\boldsymbol{e}_\ell) \cdot f(\boldsymbol{e}_\ell)$$

The goal is

- to find a strategy that maximizes its expected value.

Building strategies using the models

For all terminal nodes $\ell \in \mathcal{L}(\boldsymbol{s})$ of a strategy $\boldsymbol{s}$ we also define:

- an evaluation function $f : \cup_{\boldsymbol{s} \in \mathcal{S}} \mathcal{L}(\boldsymbol{s}) \mapsto \mathbb{R}$,
- The evaluation function can be, e.g., the information gain.
- Information gain in a node n of a strategy
  $IG(\boldsymbol{e}_n) = H(P(\boldsymbol{S})) - H(P(\boldsymbol{S} \mid \boldsymbol{e}_n))$

For each strategy we can compute:

- expected value of the strategy:

$$E_f(\boldsymbol{s}) \quad = \quad \sum_{\ell \in \mathcal{L}(\boldsymbol{s})} P(\boldsymbol{e}_\ell) \cdot f(\boldsymbol{e}_\ell)$$

The goal is

- to find a strategy that maximizes its expected value.
- Specifically, we will maximize the expected information gain,

Building strategies using the models

For all terminal nodes $\ell \in \mathcal{L}(\boldsymbol{s})$ of a strategy $\boldsymbol{s}$ we also define:

- an evaluation function $f : \cup_{\boldsymbol{s} \in \mathcal{S}} \mathcal{L}(\boldsymbol{s}) \mapsto \mathbb{R}$,
- The evaluation function can be, e.g., the information gain.
- Information gain in a node n of a strategy
  $IG(\boldsymbol{e}_n) = H(P(\boldsymbol{S})) - H(P(\boldsymbol{S} \mid \boldsymbol{e}_n))$

For each strategy we can compute:

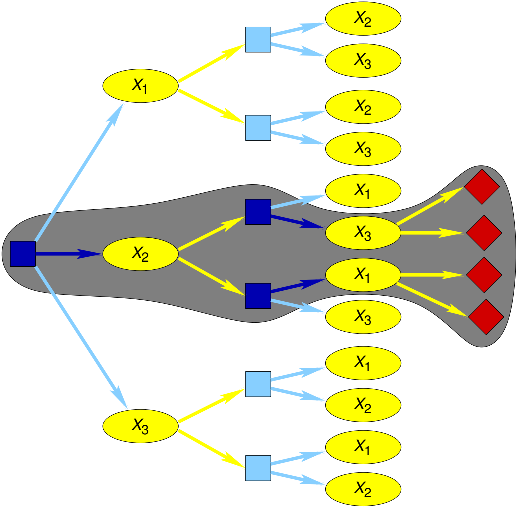
- expected value of the strategy:

$$E_f(\boldsymbol{s}) = \sum_{\ell \in \mathcal{L}(\boldsymbol{s})} P(\boldsymbol{e}_\ell) \cdot f(\boldsymbol{e}_\ell)$$

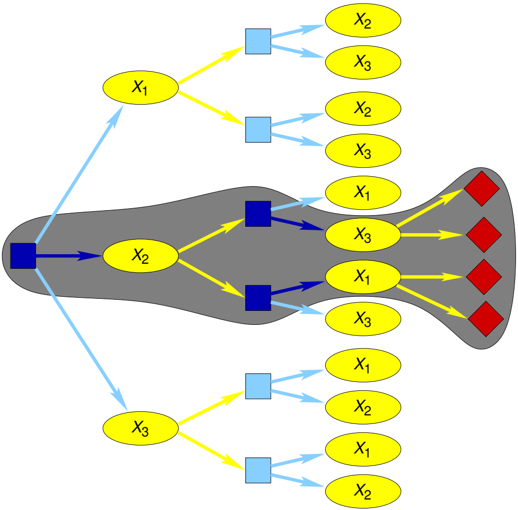
The goal is

- to find a strategy that maximizes its expected value.
- Specifically, we will maximize the expected information gain,
- which corresponds to minimization of expected entropy.

Space of tests of length 2 of 3 possible questions

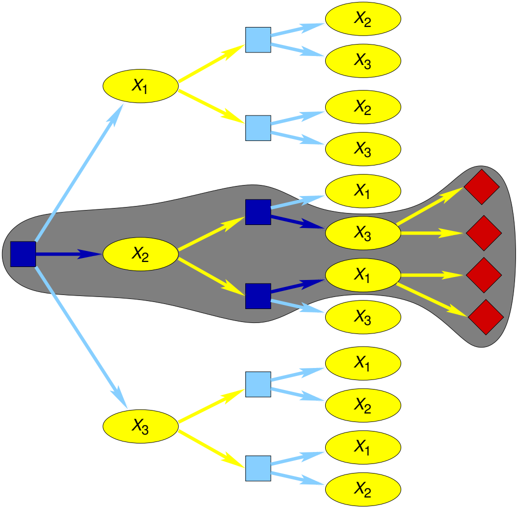


Space of tests of length 2 of 3 possible questions



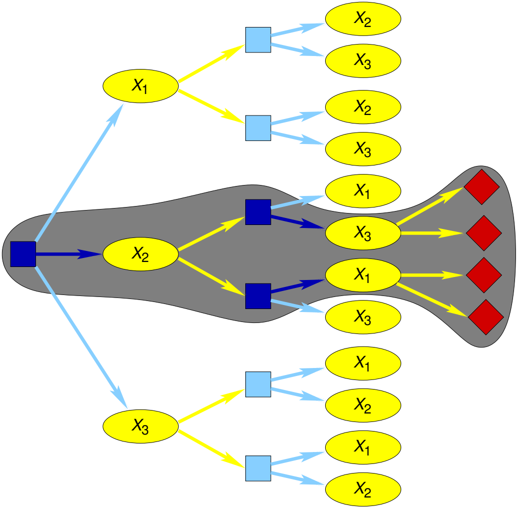
- **Entropy** in node n:
  $H(\boldsymbol{e}_n) = H(P(\boldsymbol{S} \mid \boldsymbol{e}_n))$

Space of tests of length 2 of 3 possible questions



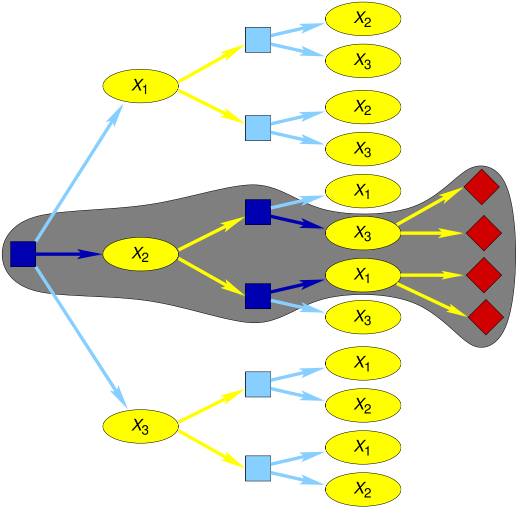
- **Entropy** in node n:
  $H(\boldsymbol{e}_n) = H(P(\boldsymbol{S} \mid \boldsymbol{e}_n))$
- **Expected entropy** of a test $\boldsymbol{t}$
  $E_H(\boldsymbol{t}) = \sum_{\ell \in \mathcal{L}(\boldsymbol{t})} P(\boldsymbol{e}_\ell) \cdot H(\boldsymbol{e}_\ell)$

Space of tests of length 2 of 3 possible questions



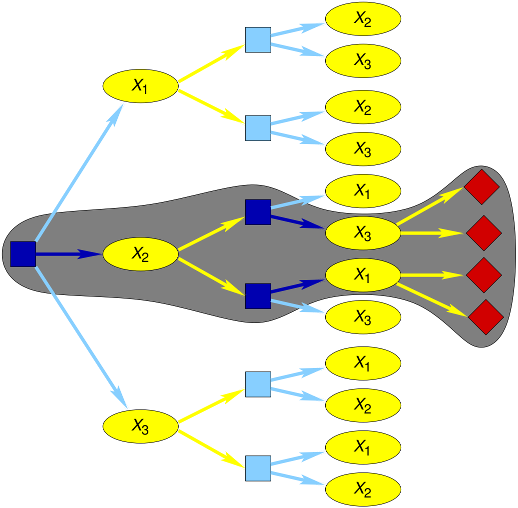
- **Entropy** in node n: $H(\boldsymbol{e}_n) = H(P(\boldsymbol{S} \mid \boldsymbol{e}_n))$
- **Expected entropy** of a test $\boldsymbol{t}$ $E_H(\boldsymbol{t}) = \sum_{\ell \in \mathcal{L}(\boldsymbol{t})} P(\boldsymbol{e}_\ell) \cdot H(\boldsymbol{e}_\ell)$
- Let $\mathcal{T}$ be the set of all possible tests (e.g. of a given length).

Space of tests of length 2 of 3 possible questions



- **Entropy** in node n: $H(\boldsymbol{e}_n) = H(P(\boldsymbol{S} \mid \boldsymbol{e}_n))$
- **Expected entropy** of a test $\boldsymbol{t}$ $E_H(\boldsymbol{t}) = \sum_{\ell \in \mathcal{L}(\boldsymbol{t})} P(\boldsymbol{e}_\ell) \cdot H(\boldsymbol{e}_\ell)$
- Let $\mathcal{T}$ be the set of all possible tests (e.g. of a given length).
- A **test $t^\star$ is optimal** iff $\boldsymbol{t}^\star = \arg\min_{\boldsymbol{t} \in \mathcal{T}} E_H(\boldsymbol{t})$.

Space of tests of length 2 of 3 possible questions



- **Entropy** in node n:
  $H(\boldsymbol{e}_n) = H(P(\boldsymbol{S} \mid \boldsymbol{e}_n))$
- **Expected entropy** of a test $\boldsymbol{t}$
  $E_H(\boldsymbol{t}) = \sum_{\ell \in \mathcal{L}(\boldsymbol{t})} P(\boldsymbol{e}_\ell) \cdot H(\boldsymbol{e}_\ell)$
- Let $\mathcal{T}$ be the set of all possible tests (e.g. of a given length).
- A **test $t^\star$ is optimal** iff
  $\boldsymbol{t}^\star = \arg\min_{\boldsymbol{t} \in \mathcal{T}} E_H(\boldsymbol{t})$.
- **Myopically optimal test**:
  in each step a we select a question that minimizes expected entropy of the test that would terminate after this question (one step look ahead).

Adaptive test of basic operations with fractions

Examples of tasks:

$$
\begin{array}{llcl}
T_1: & \left(\frac{3}{4} \cdot \frac{5}{6}\right) - \frac{1}{8} & = & \frac{15}{24} - \frac{1}{8} = \frac{5}{8} - \frac{1}{8} = \frac{4}{8} = \frac{1}{2} \\
T_2: & \frac{1}{6} + \frac{1}{12} & = & \frac{2}{12} + \frac{1}{12} = \frac{3}{12} = \frac{1}{4} \\
T_3: & \frac{1}{4} \cdot 1\frac{1}{2} & = & \frac{1}{4} \cdot \frac{3}{2} = \frac{3}{8} \\
T_4: & \left(\frac{1}{2} \cdot \frac{1}{2}\right) \cdot \left(\frac{1}{3} + \frac{1}{3}\right) & = & \frac{1}{4} \cdot \frac{2}{3} = \frac{2}{12} = \frac{1}{6} \quad .
\end{array}
$$

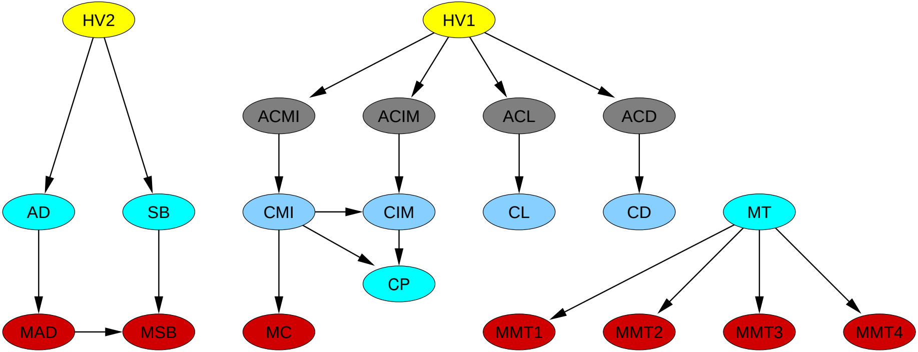
Elementary and operational skills

CP	Comparison (common numerator or denominator)	$\frac{1}{2} > \frac{1}{3}, \quad \frac{2}{3} > \frac{1}{3}$
AD	Addition (comm. denom.)	$\frac{1}{7} + \frac{2}{7} = \frac{1+2}{7} = \frac{3}{7}$
SB	Subtract. (comm. denom.)	$\frac{2}{5} - \frac{1}{5} = \frac{2-1}{5} = \frac{1}{5}$
MT	Multiplication	$\frac{1}{2} \cdot \frac{3}{5} = \frac{3}{10}$
CD	Common denominator	$\left(\frac{1}{2}, \frac{2}{3}\right) = \left(\frac{3}{6}, \frac{4}{6}\right)$
CL	Canceling out	$\frac{4}{6} = \frac{2 \cdot 2}{2 \cdot 3} = \frac{2}{3}$
CIM	Conv. to mixed numbers	$\frac{7}{2} = \frac{3 \cdot 2 + 1}{2} = 3\frac{1}{2}$
CMI	Conv. to improp. fractions	$3\frac{1}{2} = \frac{3 \cdot 2 + 1}{2} = \frac{7}{2}$

Misconceptions

Label	Description	Occurrence
MAD	$\frac{a}{b} + \frac{c}{d} = \frac{a+c}{b+d}$	14.8%
MSB	$\frac{a}{b} - \frac{c}{d} = \frac{a-c}{b-d}$	9.4%
MMT1	$\frac{a}{b} \cdot \frac{c}{b} = \frac{a \cdot c}{b}$	14.1%
MMT2	$\frac{a}{b} \cdot \frac{c}{b} = \frac{a+c}{b \cdot b}$	8.1%
MMT3	$\frac{a}{b} \cdot \frac{c}{d} = \frac{a \cdot d}{b \cdot c}$	15.4%
MMT4	$\frac{a}{b} \cdot \frac{c}{d} = \frac{a \cdot c}{b+d}$	8.1%
MC	$a\frac{b}{c} = \frac{a \cdot b}{c}$	4.0%

Student model



Evidence model for task $T1$

$$\left(\frac{3}{4} \cdot \frac{5}{6}\right) - \frac{1}{8} = \frac{15}{24} - \frac{1}{8} = \frac{5}{8} - \frac{1}{8} = \frac{4}{8} = \frac{1}{2}$$

Evidence model for task $T1$

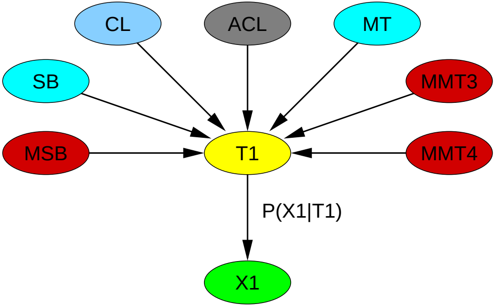
$$\left(\frac{3}{4} \cdot \frac{5}{6}\right) - \frac{1}{8} = \frac{15}{24} - \frac{1}{8} = \frac{5}{8} - \frac{1}{8} = \frac{4}{8} = \frac{1}{2}$$

$$T1 \quad \Leftrightarrow \quad MT \,\&\, CL \,\&\, ACL \,\&\, SB \,\&\, \neg MMT3 \,\&\, \neg MMT4 \,\&\, \neg MSB$$

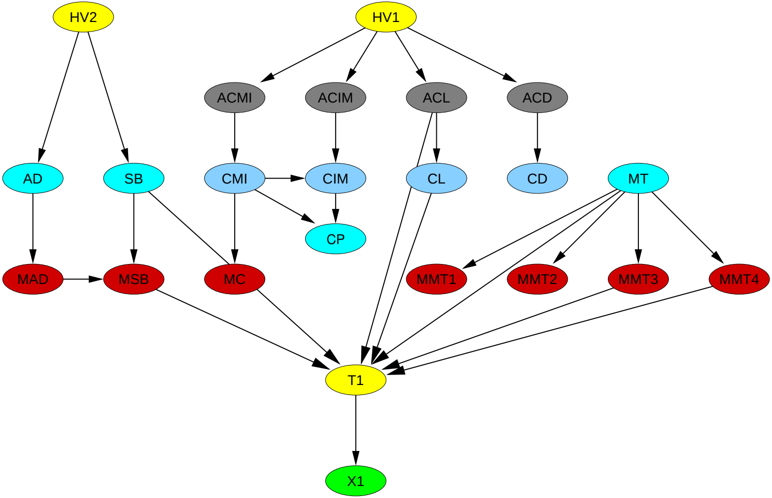
Evidence model for task $T1$

$$\left(\frac{3}{4}\cdot\frac{5}{6}\right)-\frac{1}{8}=\frac{15}{24}-\frac{1}{8}=\frac{5}{8}-\frac{1}{8}=\frac{4}{8}=\frac{1}{2}$$

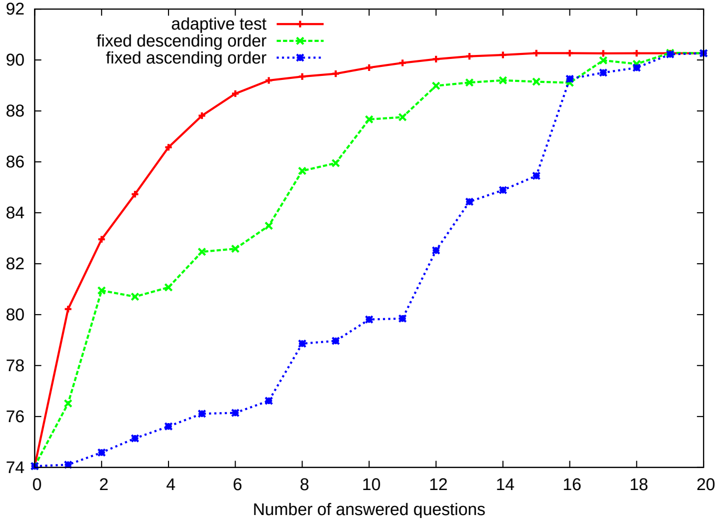
$$T1 \quad \Leftrightarrow \quad MT \ \& \ CL \ \& \ ACL \ \& \ SB \ \& \ \neg MMT3 \ \& \ \neg MMT4 \ \& \ \neg MSB$$



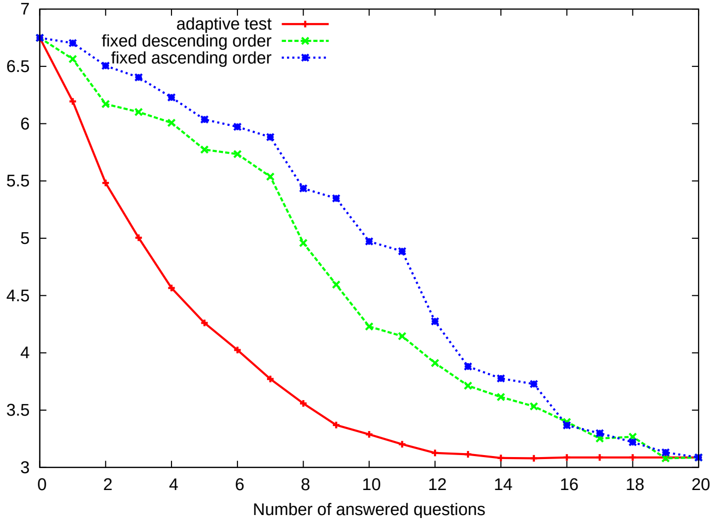
Evidence model for task $T1$ connected with the student model



Skill Prediction Quality



Entropy



7
6.5
6
5.5
5
4.5
4
3.5
3
0
2
4
6
8
10
12
14
16
18
20
adaptive test
fixed descending order
fixed ascending order
Number of answered questions

Conclusions and References

- We have shown how Bayesian networks can be used for the construction of adaptive tests in CAT.

Conclusions and References

- We have shown how Bayesian networks can be used for the construction of adaptive tests in CAT.
- Adaptive tests can substantially reduce the number of questions we need to ask a tested student.

Conclusions and References

- We have shown how Bayesian networks can be used for the construction of adaptive tests in CAT.
- Adaptive tests can substantially reduce the number of questions we need to ask a tested student.

References:

Conclusions and References

- We have shown how Bayesian networks can be used for the construction of adaptive tests in CAT.
- Adaptive tests can substantially reduce the number of questions we need to ask a tested student.

References:

- Howard Wainer, David Thissen, and Robert J. Mislevy. Computerized Adaptive Testing: A Primer. Mahwah, N.J., Lawrence Erlbaum Associates, Second edition, 2000.

Conclusions and References

- We have shown how Bayesian networks can be used for the construction of adaptive tests in CAT.
- Adaptive tests can substantially reduce the number of questions we need to ask a tested student.

References:

- Howard Wainer, David Thissen, and Robert J. Mislevy. Computerized Adaptive Testing: A Primer. Mahwah, N.J., Lawrence Erlbaum Associates, Second edition, 2000.
- Russell G. Almond and Robert J. Mislevy. Graphical models and computerized adaptive testing. Applied Psychological Measurement, Vol. 23(3), pp. 223–237, 1999.

Conclusions and References

- We have shown how Bayesian networks can be used for the construction of adaptive tests in CAT.
- Adaptive tests can substantially reduce the number of questions we need to ask a tested student.

References:

- Howard Wainer, David Thissen, and Robert J. Mislevy. Computerized Adaptive Testing: A Primer. Mahwah, N.J., Lawrence Erlbaum Associates, Second edition, 2000.
- Russell G. Almond and Robert J. Mislevy. Graphical models and computerized adaptive testing. Applied Psychological Measurement, Vol. 23(3), pp. 223–237, 1999.
- J. Vomlel: Bayesian networks in educational testing, International Journal of Uncertainty, Fuzziness and Knowledge Based Systems, Vol. 12, Supplementary Issue 1, 2004, pp. 83—100.