Bayesian Networks in Educational Assessment Tutorial

Session I: Evidence Centered Design Bayesian Networks

Duanli Yan, Diego Zapata, ETS
Russell Almond, FSU

Agenda

SESSION	TOPIC	PRESENTERS
Session 1:	Evidence Centered Design Bayesian Networks	Diego Zapata
Session 2:	Bayes Net Applications ACED: ECD in Action	Duanli Yan & Russell Almond
Session 3:	Refining Bayes Nets with Data	Russell Almond
Session 4:	Bayes Nets with R	Duanli Yan & Russell Almond

The Interplay of Design and Statistical Modeling

- Statistical models must be selected/tailored according to the needs of the assessment
- Such selection and adaptation is only meaningful in the larger context of the assessment design
- Understanding the discipline of assessment design is a necessary prerequisite for statistical modeling
- Evidence Centered Design is an assessment design framework with general applicability and utility

Test Design

- Stakeholders
- Requirements
  - Purpose of the test
  - Intended population
- Prospective Score Report
- Evidence-Centered Design
  - Claims
  - Validity
- Specifications

Evidence Centered Design

- Evidence Centered Design (ECD) provides a mechanism for
  - **Capturing and documenting information** about the structure and strength of evidentiary relationships.
  - **Coordinating the work** of test developers in authoring tasks and psychometricians in calibrating the measurement model.
  - **Documenting the scientific information** that provides the foundation for the assessment and its validity.

Evidence Centered Design

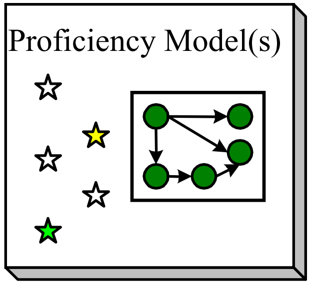
- The Evidence Centered Design process is a series of procedures which center around the questions:
  - "What can we observe about an examinee's performance which will provide evidence that the examinee has or does not have the knowledge, skills and abilities we wish to make claims about?"
  - "How can we structure situations to be able to make those observations?"
- This process results in a formal design for an assessment we call the **Conceptual Assessment Framework (CAF)**

The Initial Frame

- *Why* are we measuring?
  - What are the goals and the desires for use of this assessment?
  - Prospective Score Report
- *Who* are we measuring?
  - Who would take the assessment?
  - Who would view results and for what purpose?
- Goals of the assessment that represent the targets around which the rest of the design process is oriented

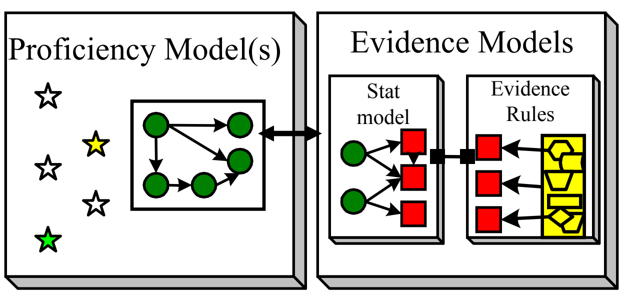
Conceptual Assessment Framework (CAF)

What we measure = Student **Proficiency** Model



Conceptual Assessment Framework (CAF)

What we measure = Student **Proficiency** Model
How we measure = **Evidence** Model

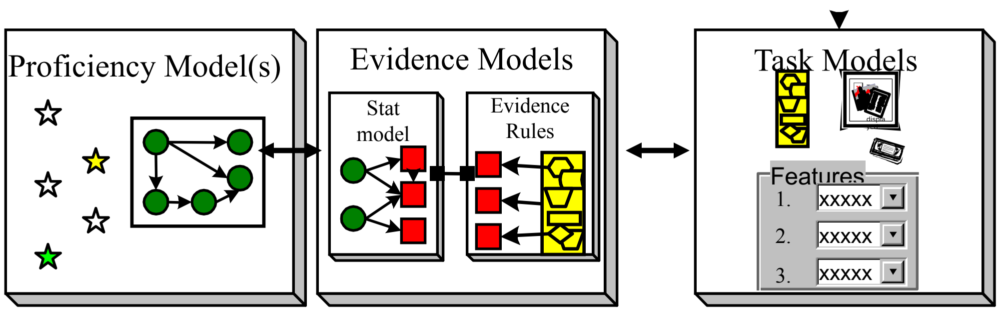


Conceptual Assessment Framework (CAF)

What we measure = Student **Proficiency** Model

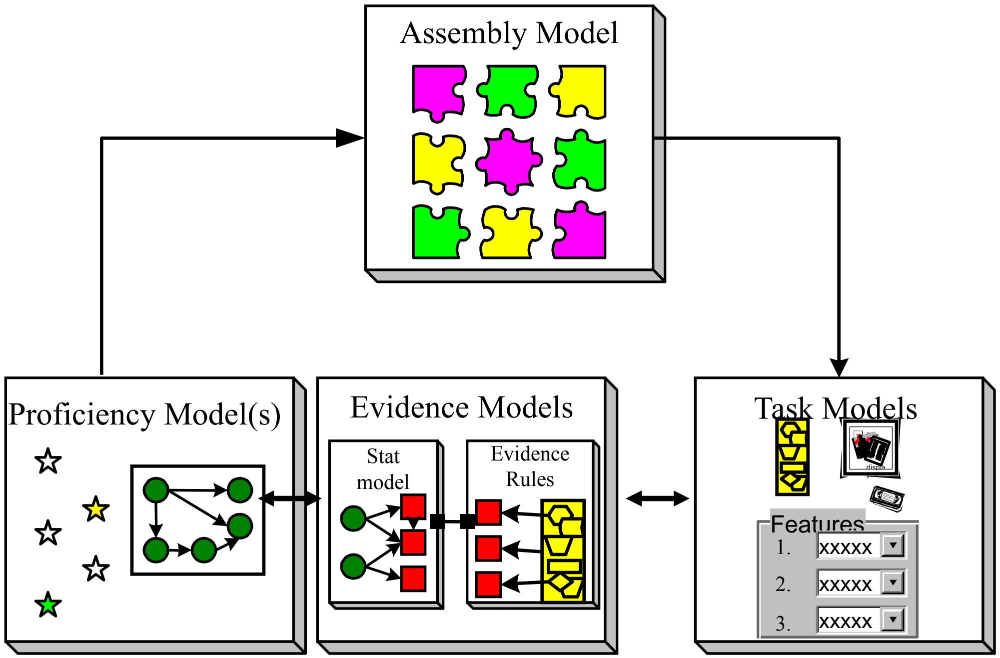
How we measure = **Evidence** Model

Where we measure = **Task** Model



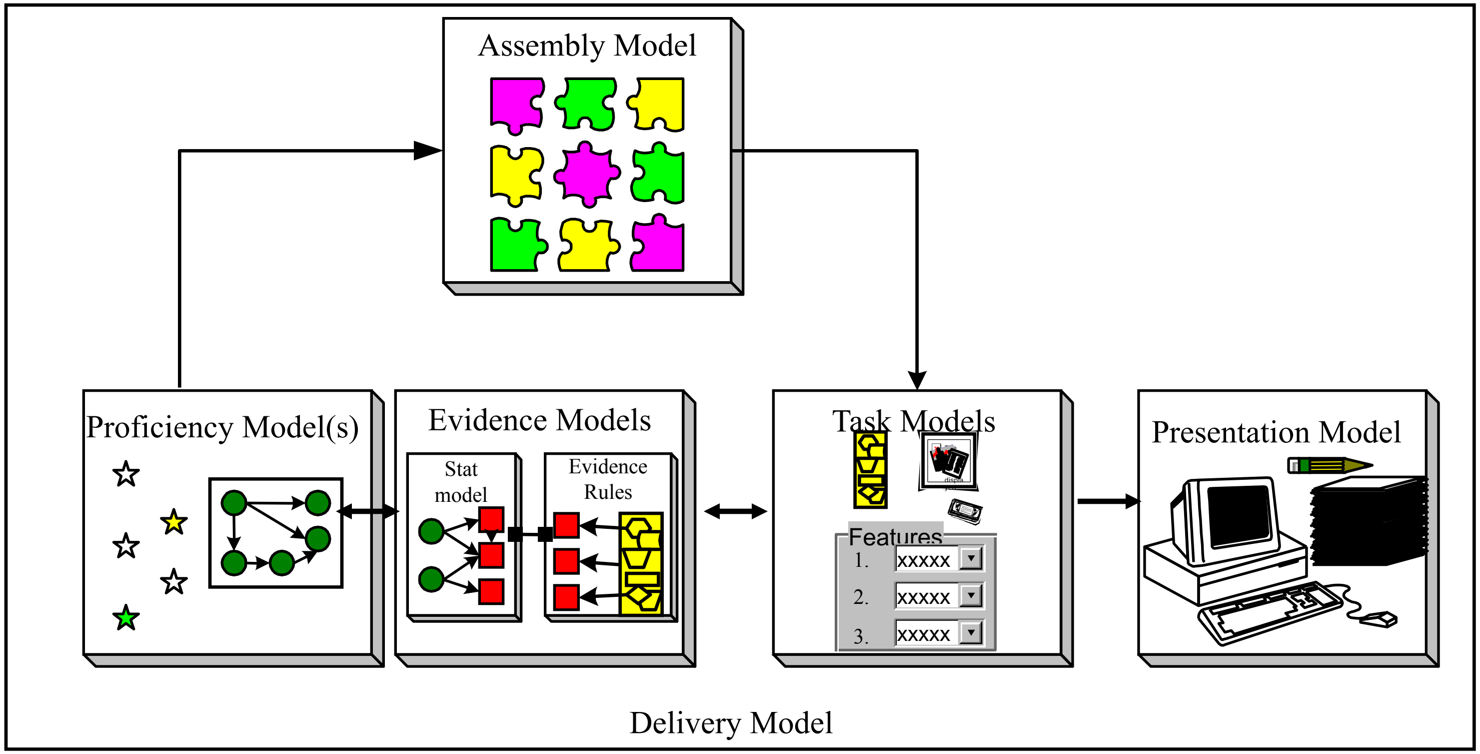
Conceptual Assessment Framework (CAF)

What we measure = Student **Proficiency** Model
How we measure = **Evidence** Model
Where we measure = **Task** Model
How Much we measure = **Assembly** Model



Conceptual Assessment Framework (CAF)

What we measure = Student **Proficiency** Model
How we measure = **Evidence** Model
Where we measure = **Task** Model
How Much we measure = **Assembly** Model
Customization = **Presentation & Delivery** Models



Activity 1: Driver's License Exam

- Redesign the driver's licensure exam
- Write down several claims you would like to make about people who receive a driver's license
- Group your claims into several proficiency variables related to the driver's test
- Do the claims hold for high, medium or low values of those variables?
- Use Netica as a drawing tool and add your variables

Activity 1 (cont)

- List a bunch of activities that you may want prospective drivers to do in their exam
- What is environment of the task
- What are manipulable features of the task?
- Pick one of the tasks you created and build an evidence model for it.
- What are some observable outcomes? their possible values?
- Which proficiencies do they measure?

Activity 1 (cont)

- Think a bit about putting this driver's test together
- How many tasks do we need of what types?
- How much time will be spent in written tests? On the road? In simulators?
- How do we verify the identity of applicants?

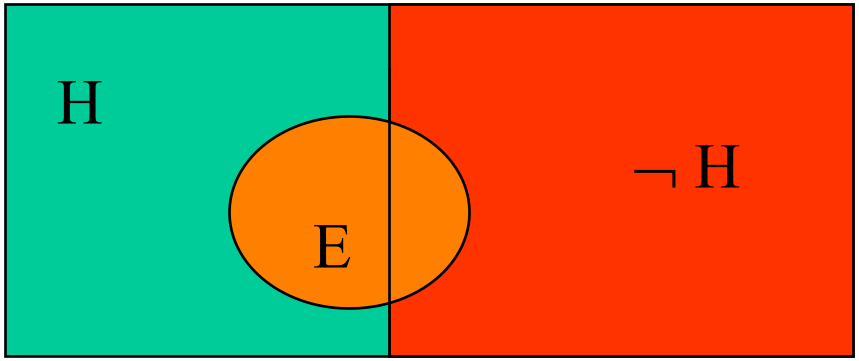
ECD $\rightarrow$ Bayes Nets

- Represent Qualitative ECD argument with a graph (Domain Modeling) (Session I)
- Turn graphical structure into probability distribution over proficiency variables and observable outcomes (Bayes net; Session I)
- Perform inference (scoring) using that Bayes net (Session II)
- Express probabilities in terms of unknown parameters -- learn parameters (Session III)
- Refine model based on how well it fits data (Session IV)

Cup and Cap notation

- In probability theory, events are sets (sets of balls in the urn).
- Let *A* and *B* be two events
- Either *A or B* occurs
  - Corresponds to *union* of sets
  - $A \cup B$
- Both *A and B* occur
  - Corresponds to *intersection* of sets
  - $A \cap B$
  - Sometimes $Pr(A,B)$
- *Not A* – the balls in the urn that are not in event *A*
  - $\neg A$
  - $Pr(\neg A) = 1\text{-}Pr(A)$

Conditional Probability



- Definition

$$\Pr(E|H) = \frac{\Pr(E \cap H)}{P(H)}$$

- Law of Total Probability

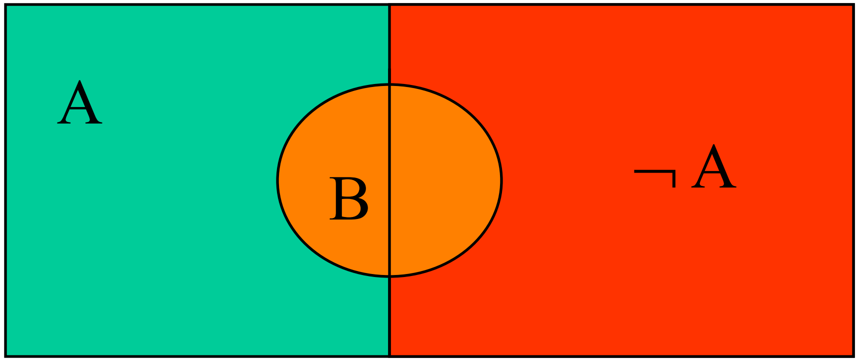
$$\Pr(E) = \Pr(E|H)\Pr(H) + \Pr(E|\neg H)\Pr(\neg H)$$

Bayes Theorem

$$
\begin{aligned}
\Pr(H|E) &= \frac{\Pr(E|H)\Pr(H)}{\Pr(E)} \\
&= \frac{\Pr(E|H)\Pr(H)}{\Pr(E|H)\Pr(H) + \Pr(E|\neg H)\Pr(\neg H)}
\end{aligned}
$$

- Prior $\Pr(H)$
- Likelihood $\Pr(E|H)$
- Posterior $\Pr(H|E)$

Independence



$$
\begin{aligned}
\Pr(B) &= \Pr(B|A) &&= \Pr(B|\neg A) \\
\Pr(A) &= \Pr(A|B) &&= \Pr(A|\neg B) \\
\Pr(A \cap B) &= \Pr(B|A)\Pr(B) &&= \Pr(A)\Pr(B)
\end{aligned}
$$

- A provides no information about B

Accident Proneness (Feller, 1968)

- Driving Skill: 5/6 Normal, 1/6 Accident Prone
- Probability of an accident in a given year
  - 1/100 for Normal drivers
  - 1/10 for Accident prone drivers
- Accidents happen independently in each year
- What is the probability a randomly chosen driver will have an accident in Year 1?
- Given a driver had an accident in Year 1, what is probability of accident in Year 2?

Accident Proneness (cont)

- What is the probability a randomly chosen driver will have an accident in Year 1? Year 2?

$$\begin{aligned} P(A_i) &= P(A_i|N)P(N) + P(A_i|\overline{N})P(\overline{N}) \\ &= \frac{.05}{6} + \frac{.1}{6} = \frac{.15}{6} = .025\ . \end{aligned}$$

Accident Proneness (cont)

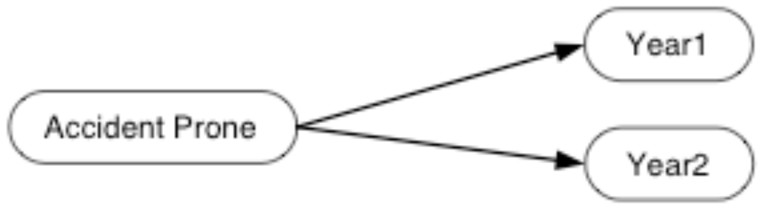
- Given a driver had an accident in Year 1, what is probability of accident in Year 2?

$$\begin{aligned}
P(A_1 \cap A_2) &= P(A_1 \cap A_2|N)P(N) + P(A_1 \cap A_2|\overline{N})P(\overline{N}) \\
&= P(A_1|N)P(A_2|N)P(N) + P(A_1|\overline{N})P(A_2|\overline{N})P(\overline{N}) \\
&= .01 \times .01 \times \frac{5}{6} + .1 \times .1 \times \frac{1}{6} \\
&= \frac{.0005}{6} + \frac{.01}{6} = \frac{.0105}{6} = .00175 \ .
\end{aligned}$$

Note that

$$P(A_2|A_1) = \frac{P(A_1 \cap A_2)}{P(A_2)} = \frac{.00175}{.025} = .07 \ .$$

Conditional Independence

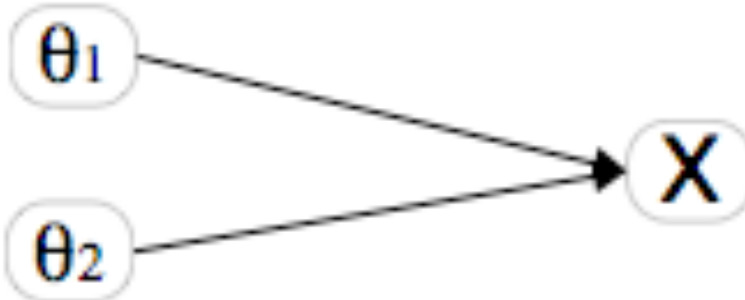


- Years are *conditionally independent* given driving skill

$$p(Y_1, Y_2|S) = p(Y_1|S)p(Y_2|S)$$

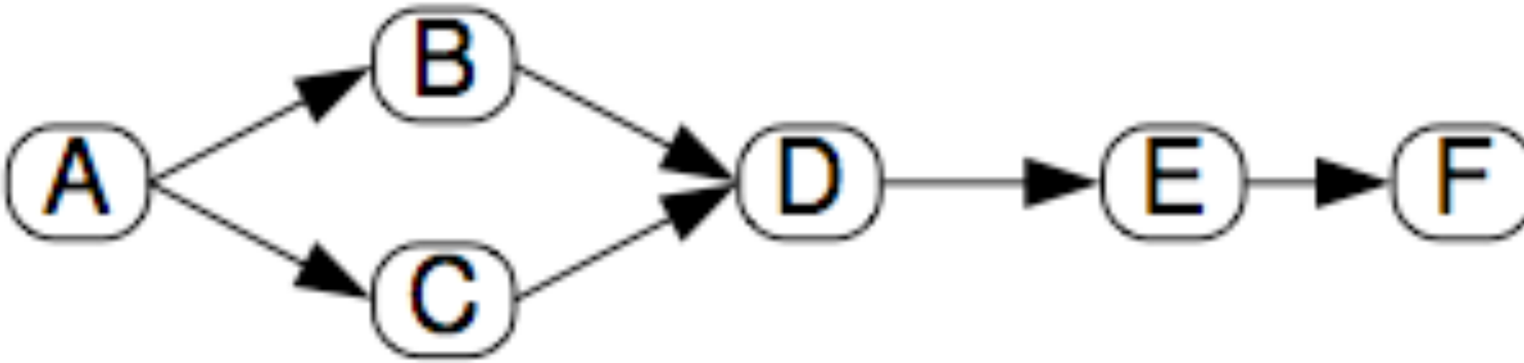
- Years are *marginally dependent*
- Separation in graph tells the story

Competing Explanations



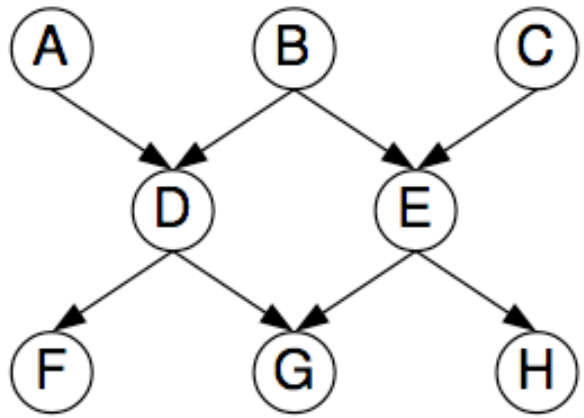
- Skill 1 and Skill 2 are (a priori) independent in population
- Task X requires both skills (conjunctive model)
- Answer the following questions:
  - What is posterior after learning X=False, and θ_1=High?
  - What is posterior after learning X=False, and θ_2=High?
  - What is true of joint posterior of θ_1 and θ_2□□after learning X=False?

D-Separation



- For ←→, →→, and ←← edges conditioning on middle variables renders outer variables independent
- For →← (collider) edges, if middle variable (or descendent is known) then variables are dependent
- A path is *active* if collider with middle node observed, or non-collider with middle node unobserved

D-Separation Exercise



- Are *A* and *C* independent if
  1. We have observed no other variables?
     - What could we condition on to make *A* and *C* independent?
  2. We have observed *F* and *H?*
     - What else could we condition on to make *A* and *C* independent?
  3. We have observed *G*?
     - What else could we condition on to make *A* and *C* independent?

Building Up Complex Networks

- Recursive representation of probability distributions:

$$p(x_1,\ldots,x_n) = p(x_n \mid x_{n-1},\ldots,x_1)p(x_{n-1} \mid x_{n-2},\ldots,x_1)\cdots p(x_2 \mid x_1)p(x_1)$$

$$= \prod_{j=1}^{n} p(x_j \mid x_{j-1},\ldots,x_1) = \prod_{j=1}^{n} p(x_j \mid Pa(x_j)),$$

- All orderings are equally correct, but some are more beneficial because they capitalize on causal, dependence, time-order, or theoretical relationships that we posit:

  Terms simplify when there is conditional independence – in ed measurement, due to unobservable student variables.

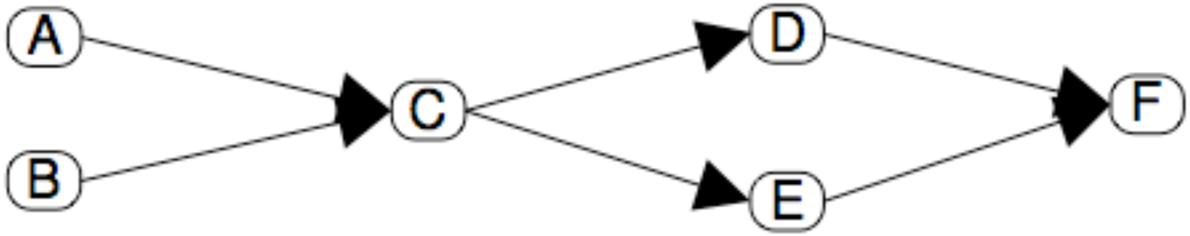
Building Up Complex Networks, cont.

- For example, in IRT, item responses are conditionally independent given θ:

$$
\begin{aligned}
&p(x_1,\ldots,x_n,\theta) \\
&= p(x_n \mid x_{n-1},\ldots,x_1,\theta)\,p(x_{n-1} \mid x_{n-2},\ldots,x_1,\theta)\cdots p(x_2 \mid x_1,\theta)\,p(x_1 \mid \theta)\,p(\theta) \\
&= p(x_n \mid \theta)\qquad p(x_{n-1} \mid \theta)\qquad \cdots p(x_2 \mid \theta)\qquad p(x_1 \mid \theta)\,p(\theta)
\end{aligned}
$$

$$
= \prod_{j=1}^{n} p(x_j \mid \theta)\,p(\theta).
$$

Bayes net



- One factor for each node in graph in recursive representation
- This factor is conditioned on parents in graph
- "Prior" nodes have no parents
- $p(A)p(B)p(C|A,B)p(D|C)p(E|C)p(F|D,E) = p(A,B,C,D,E,F)$
- Digraph must be acyclic

Activity 2: Build a Bayes Net

- Pick one of the tasks you created and build an a Bayes net in Netica:
- Proficiency variables, their possible values
- Observable variables, their possible values
- Conditional probabilities between Proficiency variables and Observable variables
- Add your observables to the proficiency model you made in Netica