Finding improvements in student models for intelligent tutoring systems via variable selection for a linear logistic test model

Brian W. Junker
Carnegie Mellon University
brian@stat.cmu.edu

Ken Koedinger
Carnegie Mellon University
koedinger@cmu.edu

Mario Trottini
Carnegie Mellon University
trottini@stat.cmu.edu

Research supported in part by NSF grant #LIS-97.20354 to Marsha C. Lovett (PI), Joel B. Greenhouse, Brian W. Junker, Ken Koedinger, and Robert E. Kass.

A Progress Report

Background: Computer-based Cognitive Tutors

- A class of *Intelligent Tutoring Systems* (ITS)
- Support learning by doing
  - Cognitive Tutor adds to limited individual attention that teacher can provide
- Cognitive Principles of Instruction
  - Make hidden thinking processes visible
  - Build from students' prior knowledge
- Source of power: The details of the cognitive student model
  - Uncover subtleties of student learning
  - Model subtleties in a running computer simulation — *the theory has to work*

ACT-R Based Tutors

- ACT-R* incorporates both connectionist and production system features, to model human cognition.
- Longstanding R&D effort at Carnegie Mellon aimed at building cognitive tutors on top of ACT-R, in:
  - LISP
  - Algebra
  - Geometry
  - ...
  - Statistics [in development]
- We are developing a methodology, using the Geometry tutor, to be applied to the Statistics tutor.

---

*Anderson, J.R. (1993). *Rules of the mind.* Hillsdale NJ: Erlbaum.

Underlying Cognitive Theory: ACT-R

- ACT-R models cognitive processes using two types of knowledge representation.

- Declarative knowledge: things we are aware we know and can usually describe to others. (e.g. *facts*)
  - Fundamental units: chunks
  - Arranged in a partially hierarchical connectionist network.
  - “Activation” determines “recallability”; increases with use.

- Procedural knowledge: knowledge which we display in our behavior but which we are not conscious of. (e.g. *automated skills*)
  - Fundamental units: Production rules
  - If/then rules for creating or modifying chunks.
  - “Activation” of chunks and production rules determines whether this rule is selected; increases with use.

ACT-R Tutor Technology

- Student Model: Incorporates multiple strategies and typical student misconceptions

```
Strategy 1:     IF the goal is to solve a(bx+c) = d
                THEN rewrite this as     bx + c = d/a

Strategy 2:     IF the goal is to solve a(bx+c) = d
                THEN rewrite this as   abx + ac = d

Misconception: IF the goal is to solve a(bx+c) = d
                THEN rewrite this as    abx + c = d
```

- Model Tracing: Follows student through their individual approach a problem: context-sensitive instruction
- Knowledge Tracing: Assesses student's knowledge growth: individualized activity selection and pacing

Successes and Problems

- Success: Cognitive Tutors dramatically enhance student learning*
  - Controlled, full year classroom experiments replicated over 3 years in urban schools In Pittsburgh and Milwaukee
  - 50–100% better on problem solving and representation use;
  - 15–25% better on standardized tests (ITBS; SAT subset).
- Problem: NOT easy to get the details of the cognitive model right
- Solution: Data-driven improvements
  - Collect volumes of data on student learning
  - Fit reasonable approximations to the data quickly to sift through many alternative models

---

*Koedinger, Anderson, Hadley, & Mark (1995). Intelligent tutoring goes to school in the big city. In J. Greer (Ed.), *Proc. 7th World Conf. Art. Int. & Ed.* AACE, Charlottesville, NC

Project: Data-Driven Improvement of Cognitive Model

- Initial cognitive model comes from analysis of *student work*, *teachers* and *teaching materials*, *experts*, etc.
- But, e.g.: Rules of mathematics $\neq$ Rules of mathematical thinking
  - Rules of thinking determine <u>when</u>, not just <u>how</u>
  - Rules of thinking are induced from experience
- Content knowledge $\neq$ Pedagogical content knowledge
- Risks of “expert blindspot”

Some Predictions of ACT-R

- Local independence: At appropriate granularity, execution of different production rules is conditionally independent given person.
- Learning curves: The odds of making an error in decrease as a power function of opportunity to apply (OTA) for each rule:

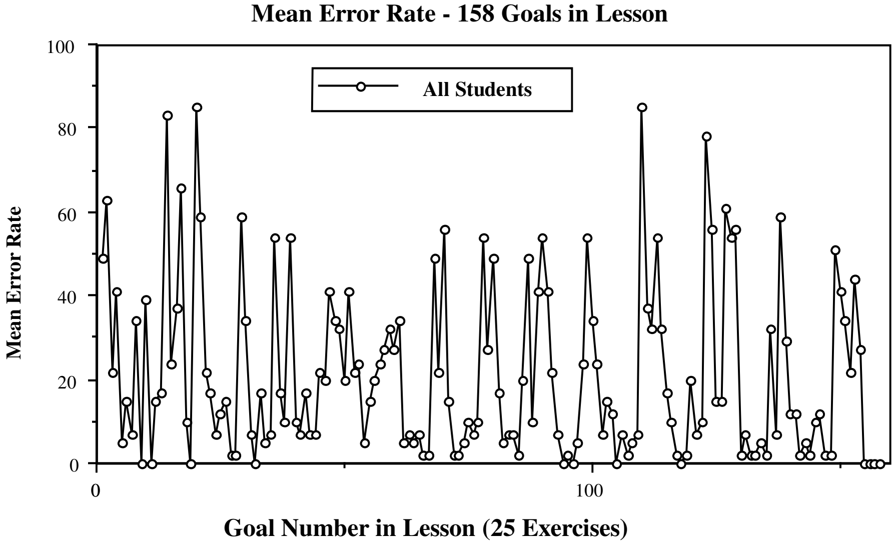
$$\frac{p}{1-p} = \alpha \cdot (OTA)^{-\beta}$$

- Individual differences: Students start at different points on the learning curve, but difficulty and rate of learning are only rule-dependent, not student-dependent or task-dependent.
- Borne out for example, in by-hand iterations of the LISP tutor*
(next two slides)

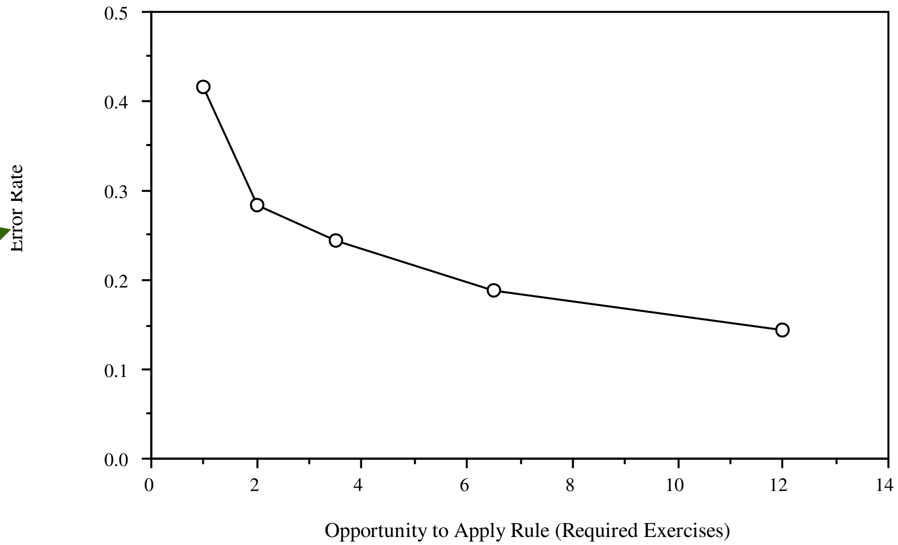
---

* Anderson, Corbett, Koedinger, & Pelletier (1995). Cognitive tutors: Lessons learned. *J. Learning Sciences, 4,* 167–207.

LISP Tutor: Production Rule Analysis

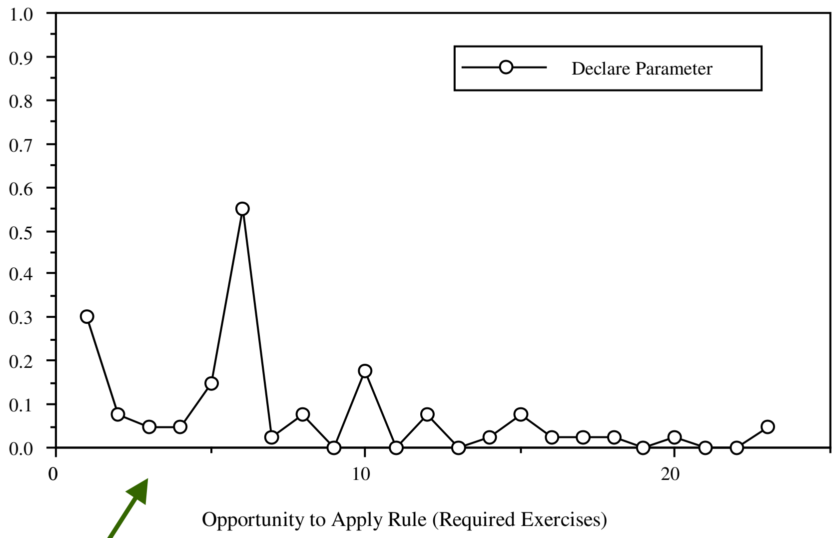


Learning?



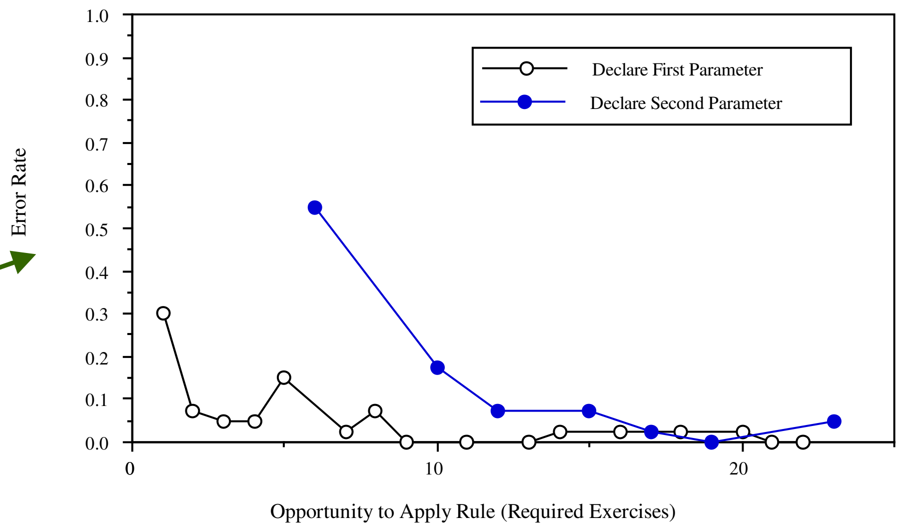
Yes! At the production rule level.

LISP Tutor: Improving Production Rules



Learning curve
is not smooth

Inspect problem
features => New
production rules
remove “blips”



Statistical Model

- The probability of student i successfully performing rule j on the t^{th} opportunity to apply:

$$\begin{aligned} \frac{p_{ijt}}{1-p_{ijt}} &= \alpha_{ijt} t_{ij}^{-\beta_{ijt}} \\ \Rightarrow\ p_{ijt} &= \frac{\exp[a_{ijt} + b_{ijt}\log(t_{ij})]}{1+\exp[a_{ijt} + b_{ijt}\log(t_{ij})]} \end{aligned}$$

  Thus* a model of learning curves with individual differences will look like the LLTM†

- We will fit error rates / learning curves from features of the cognitive model and other skill / task covariates, *<u>not reproduce cognitive model.</u>*

---

*Draney, Pirolli & Wilson (1995). A measurement model for a complex cognitive skill. In Nichols, et al. (eds.) *Cognitively diagnostic assessment.* Hillsdale, NJ: Erlbaum.

†Fischer, (1997). Unidimensional linear logistic Rasch models. In van der Linden & Hambleton (Eds.) *Handbook of modern IRT.* New York: Springer-Verlag.

Parametrization and Interpretation

- We reparametrize the model as follows:

$$\text{logit} p_{ijt} = \theta_i + \alpha_j + \beta_j \log(t_{ij})$$

  - θ_i models individual differences at the beginning of tutoring.
  - α_j models the difficulty of rule j
  - β_j models the slope of the learning curve of rule j.

- Searching for cognitive model improvements amounts to adding and deleting
  - Covariates of rule/skill difficulty
  - Covariates of rule/skill learning rate

  that improve the fit of this model.

- Terms α_j and $\beta_j \log(t_{ij})$ may be repeated in the model for multiple difficulty and learning factors

Criteria For A "Good" Cognitive Model

- Simple
  - Fewer production rules
  - Fewer parameters in LLTM
- Accurate
  - Correct grain size of knowledge acquisition
  - Good fit of statistical model to data
- Interpretable
  - Covariates should "make sense" as difficulty factors or learning factors
  - Combining covariates with existing model elements should "make sense"

Defining a Search Space

- In the Geometry Tutor, some candidate covariate factors include:
  - Embeddedness
  - Repeatedness
  - Forward-Backward
  - Polygon, Quadrilateral, Parallelogram, Rectangle
- Operators for adding and deleting covariates include
  - `Split (Skill, Factor) -> NewSkill`
  - `Add (Skill, Factor) -> Skill + Hidden-Skill`
  - `Merge (Skill, Factor) -> NewSkill`
  - Others: R-Split, Partial-Split, Partial-Add, Partial-Merge

Operator Examples

- `Split (Skill, Factor) -> NewSkill`

Problem	Skill	OTA	Factor	→	Problem	NewSkill	OTA
p1	PARALLELOGRAM-AREA	1	Alone		p1	PARALLELOGRAM-AREA-Alone	1
p1	CIRCLE-AREA	1	Embedded		p1	CIRCLE-AREA-Embedded	1
p1	CIRCLE-CIRCUMF	1	Alone		p1	CIRCLE-CIRCUMF-Alone	1
p2	CIRCLE-AREA	2	Alone		p2	CIRCLE-AREA-Alone	1
p2	CIRCLE-AREA	3	Embedded		p2	CIRCLE-AREA-Embedded	2
p2	CIRCLE-CIRCUMF	2	Embedded		p2	CIRCLE-CIRCUMF-Embedded	2

- `Split (Skill, Factor) -> NewSkill`
  - Construct `NewSkill` = `Skill` × `Factor` interaction
  - Recalculate OTA's $t'_{ij'}$ for `NewSkill`
  - Replace old $\alpha_j + \beta_j \log(t_{ij})$ terms with new $\alpha'_{j'} + \beta'_{j'} \log(t'_{ij'})$ terms
- `Add (Skill, Factor) -> Skill + Hidden-Skill`
  - <u>Difficulty Factor</u>: Add difficulty terms $\gamma_{k(j)}$ for levels k of `Factor`.
  - <u>Learning Factor</u>:
    * Compute OTA's $t_{ik(j)}$ for `Factor` as a skill
    * Add terms $\alpha_{k(j)} + \beta_{k(j)} \log(t_{ik(j)})$ to the model.

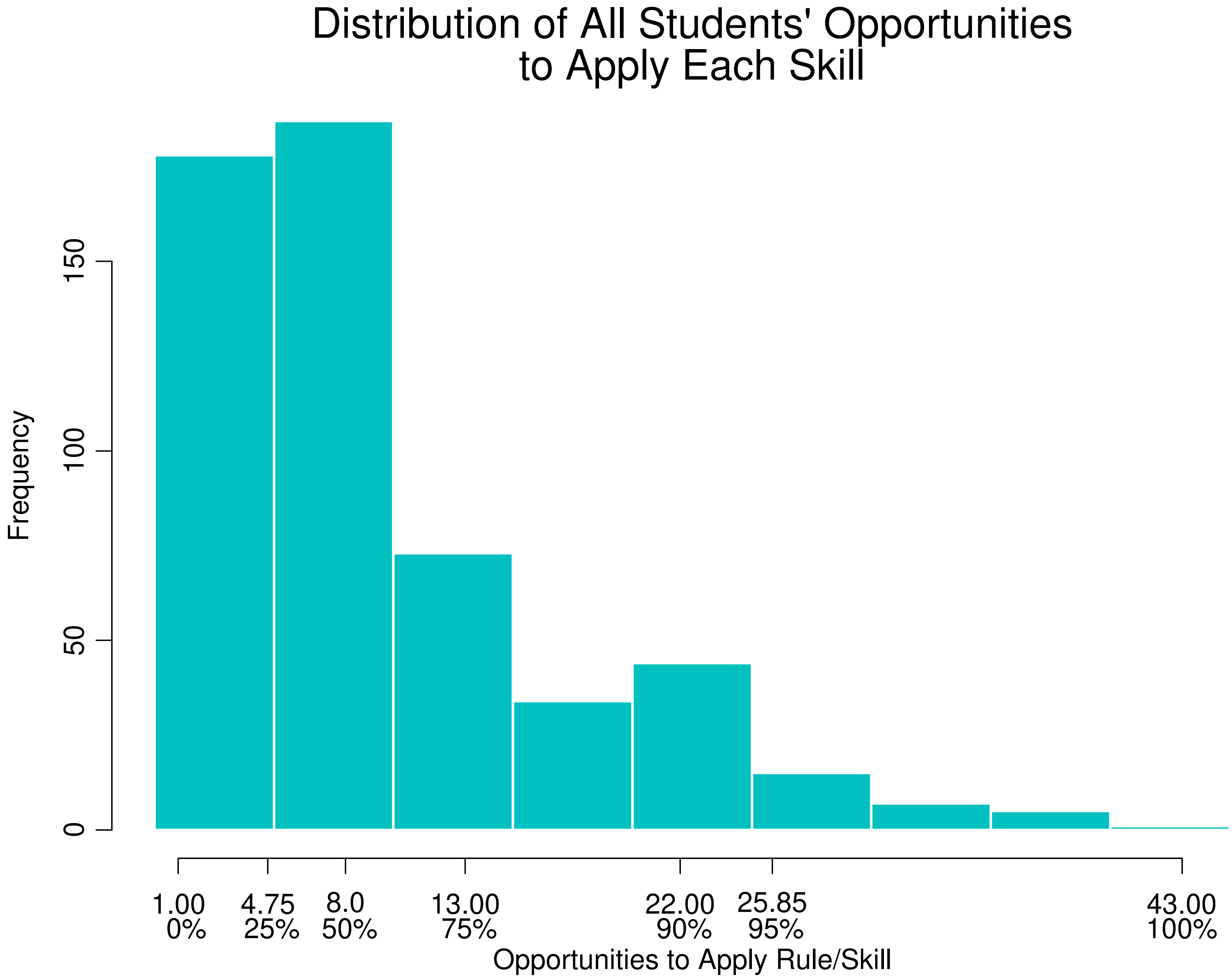
Example Using Geometry Tutor Data

- 59 Students
- 15 Skills (Production Rules)
- 5431 Skill Opportunities; 92 per student on average
- Implemented model-search (DFS) / variable-building / model-fitting (JML) in XLISP-STAT
- Compared models using BIC (Schwarz criterion*)

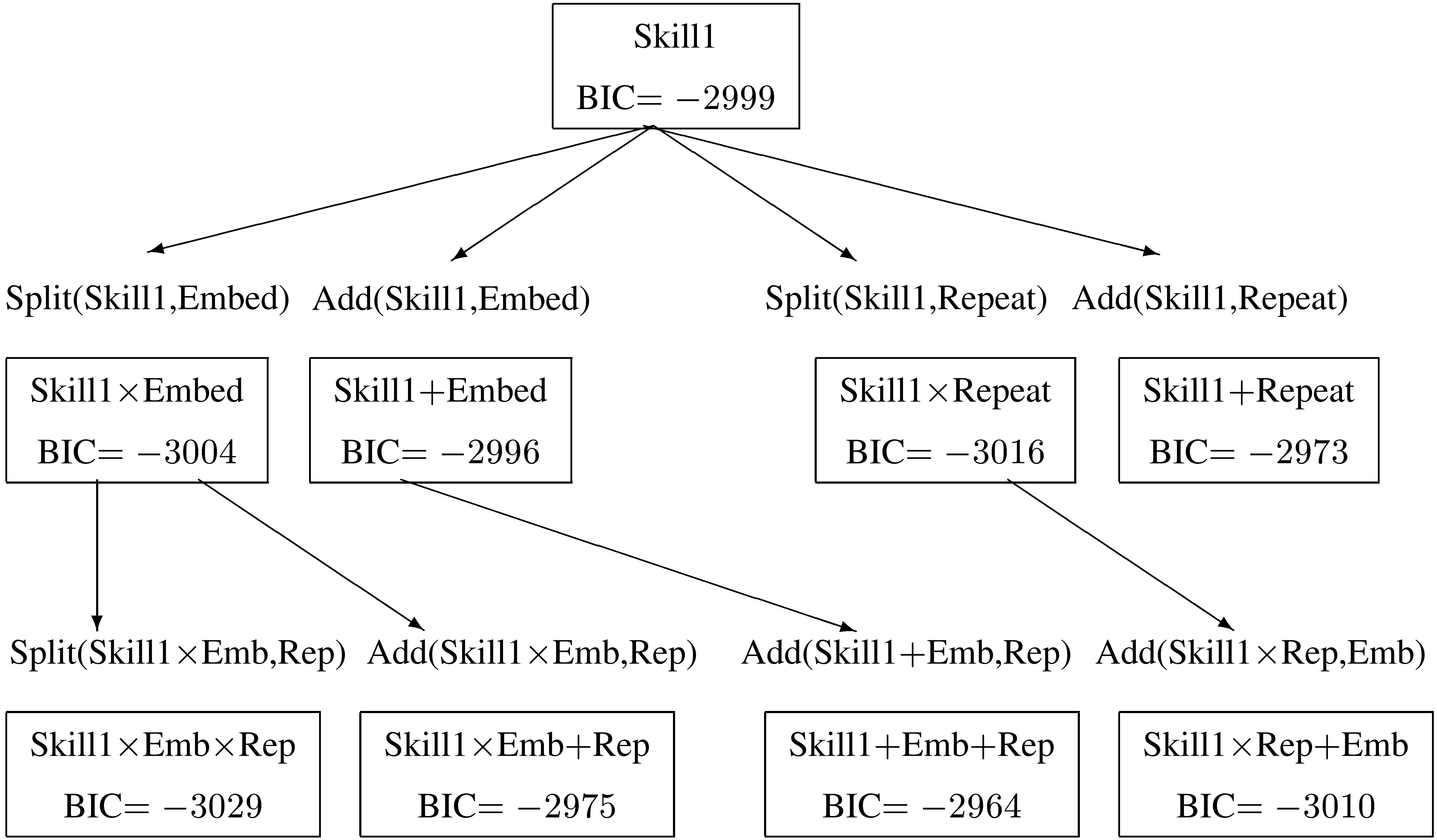
$$-2\log(\text{likelihood}) + k\log(n)$$

---

*e.g. Kass & Raftery (1995). Bayes factors. *JASA, 90,* 773–795.


Distribution of All Students' Opportunities
to Apply Each Skill
Frequency
0
50
100
150
1.00
0%
4.75
25%
8.0
50%
13.00
75%
22.00
90%
25.85
95%
43.00
100%
Opportunities to Apply Rule/Skill

Sample Model Space Search

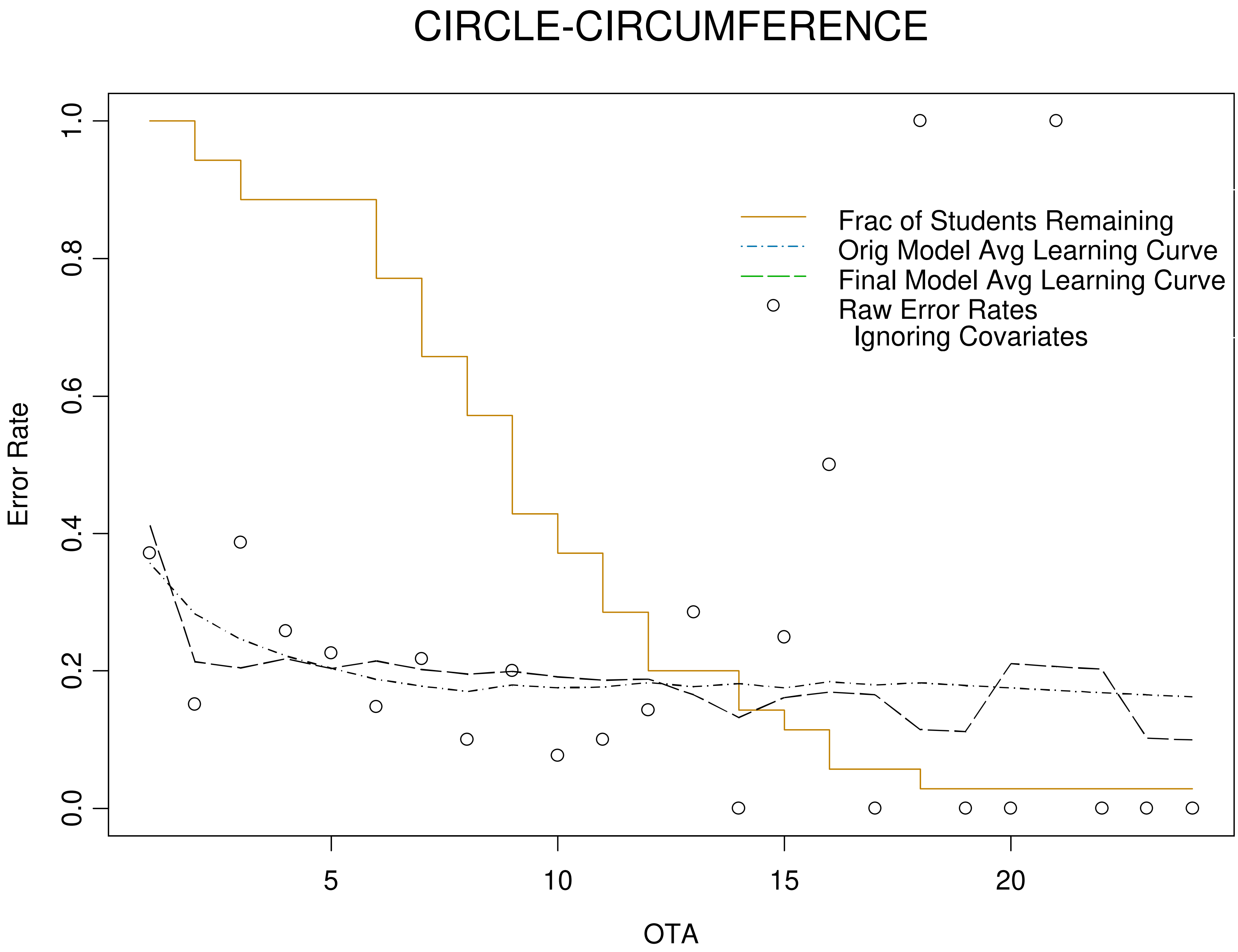


Some Preliminary Conclusions

- So far we have “proof of concept”
- Statistical analysis can reveal *hidden skills* and *hidden difficulty factors* not apparent through cognitive analysis
- What to do with them:
  - New problems to support acquiring them
  - New interfaces to make them “visible”
  - New hint messages to cue learners to them

Additional Complexities

- Drop Out
  - From 4 to 254 observations per student
  - Tutor drops student as each skill is mastered
  - Currently treating dropout as MCAR; discounting by sample size
  - Simple imputation: all-correct after dropout
  - Better imputation: use tutor's knowledge-tracing model
- Order and Gap Times
  - Students encounter opportunities to apply skills in different order
  - Gaps between OTA's from under a minute to several days
  - Our LLTM doesn't account for this


CIRCLE-CIRCUMFERENCE
Frac of Students Remaining
Orig Model Avg Learning Curve
Final Model Avg Learning Curve
Raw Error Rates
Ignoring Covariates
Error Rate
OTA
0.0
0.2
0.4
0.6
0.8
1.0
5
10
15
20

Future Work

- Model and Search Improvements:
  - Better fitting: MML, CML, MCMC
  - DFS: recognizing equivalent models
  - Dropout and gap times
  - Speed: Current example 3 hours
- Confirmation:
  - Implement other operators
  - Can we re-aquire current cognitive model from “textbook” model?
- New Domains:
  - Other parts of the Geometry Tutor
  - Statistics Tutor [in development]

References

Anderson, J. R., Corbett, A. T., Koedinger, K. R., and Pelletier, R. (1995). Cognitive tutors: lessons learned. *J. Learning Sciences, 4(2),* 167–207.

Corbett, A. T., Anderson, J. R. and O'Brien, A. T. (1995). Student modeling in the ACT programming tutor. Chapter 2 in Nichols, P. D., Chipman, S. F. and Brennan, R. L. (eds.) (1995). *Cognitively diagnostic assessment.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Draney, K. L., Pirolli, P. and Wilson, M. (1995). A measurement model for a complex cognitive skill. Chapter 5 in Nichols, P. D., Chipman, S. F. and Brennan, R. L. (eds.) (1995). *Cognitively diagnostic assessment.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Fischer, G. H. (1997). Unidimensional linear logistic Rasch models. Chapter 13, pp. 221–224 in van der Linden, W. J. and Hambleton, R. K. (eds.) (1997). *Handbook of modern item response theory.* New York: Springer-Verlag.

Lovett, M. C. and Greenhouse, J. B. (1999). Applying cognitive theory to statistics instruction. Submitted for publication. Also available at *http://lib.stat.cmu.edu/cmu-stats/tr/* as CMU Statistics Tech. Rept. #689.