Modelling Conditional Probabilities in a Complex Assessments

Robert J. Mislevy[1,2] , Deniz Senturk[3], Russell Almond[2], Lou Dibello[2], Frank Jenkins[2], Linda Steinberg[2], and Duanli Yan[2]

University of Maryland

August 13-14, 2001

[1]University of Maryland; [2]Educational Testing Service;
[3]GE; Educational Testing Service 2000 Summer Intern

Key Ideas

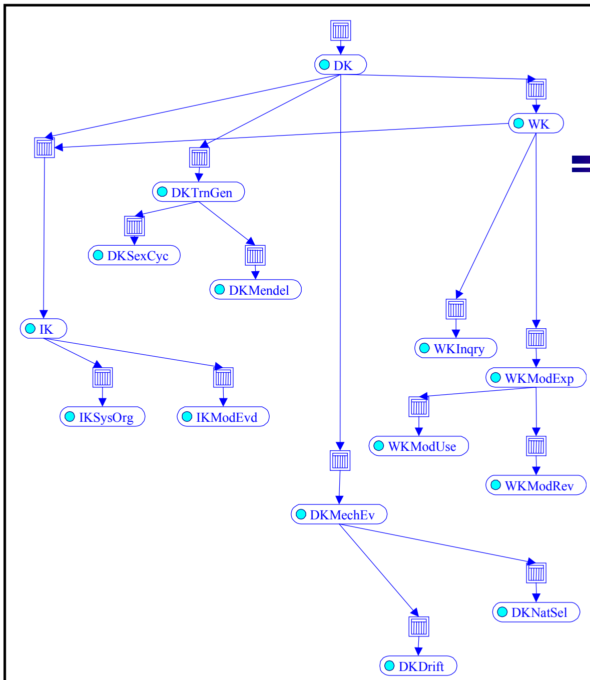
- Psychometric model for Agouti Segment 1
- Structure of model determined by structure of task & observations, and ...
- structure of task & observations determined by cognitive/substantive target of inference
- Bayesian inferential framework
- Interplay between expert / substantive information and statistical modelling
- Modular construction, from reusable pieces
- www.education.umd.edu/EDMS/mislevy/papers/ConditionalProbs.pdf

Bayesian Modelling

The first step in Bayesian analysis is setting up a full probability model, or joint probability distribution for all observable and unobservable quantities in a problem.

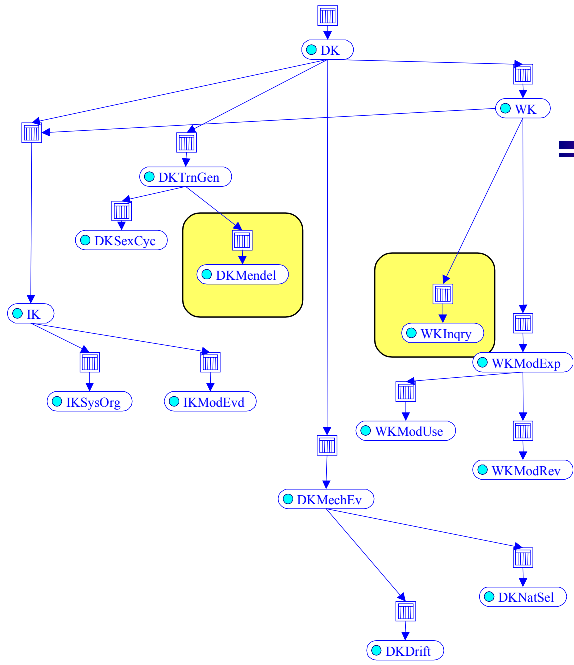
The model should be consistent with knowledge about the underlying scientific problem and the data collection process.

Gelman, Carlin, Stern, and Rubin (1995, p. 3)



Biomass SM

- Variables
- "Parents" and "Children"
- Conditional probability matrices
- Equations in a minute



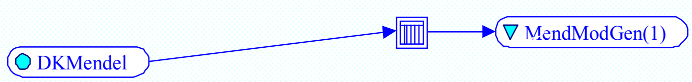
Biomass SM

From Agouti1, direct evidence about...

- DKMendel
- WKInqry

by design.

Measurement-Model fragments in EMs used in Agouti 1



DKMendel is the student-model variable that determines probabilities of response to the observable variable MendModGen(1). The conditional probability distributions are represented by [icon].

A Simple Numerical Example

(Parameters based on “Expert Opinion”)

θ	X=1 (Poor)	X=1 (Okay)	X=1 (Good)
Low	.70	.25	.05
Med	.35	.40	.25
High	.10	.40	.50

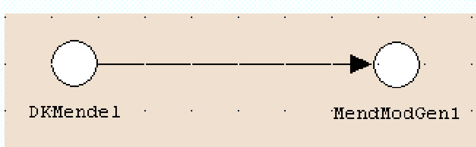
Each row is a probability distribution for the response variable X, given θ.

A Simple Numerical Example

(Parameters based on “Expert Opinion”)

θ	X=1 (Poor)	X=1 (Okay)	X=1 (Good)
Low	.70	.25	.05
Med	.35	.40	.25
High	.10	.40	.50

Each row is a probability distribution for the response variable X, given θ.



ERGO screenshot

A Simple Numerical Example

(Parameters based on “Expert Opinion”)

θ	X=1 (Poor)	X=1 (Okay)	X=1 (Good)
Low	.70	.25	.05
Med	.35	.40	.25
High	.10	.40	.50

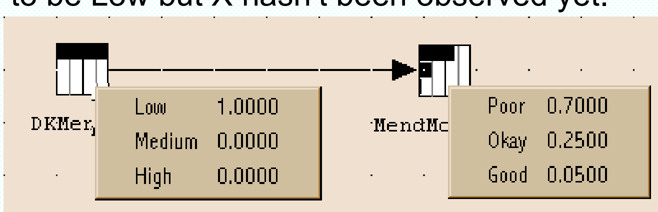
Conditional probabilities of response possibilities, if θ is known to be Low but X hasn’t been observed yet.

A Simple Numerical Example

(Parameters based on “Expert Opinion”)

θ	X=1 (Poor)	X=1 (Okay)	X=1 (Good)
Low	.70	.25	.05
Med	.35	.40	.25
High	.10	.40	.50

Conditional probabilities of response possibilities, if θ is known to be Low but X hasn’t been observed yet.

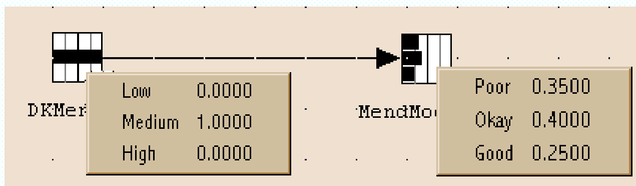


A Simple Numerical Example

(Parameters based on "Expert Opinion")

θ	X=1 (Poor)	X=1 (Okay)	X=1 (Good)
Low	.70	.25	.05
Med	.35	.40	.25
High	.10	.40	.50

Conditional probabilities of response possibilities, if θ is known to be Medium but X hasn't been observed yet.



A Simple Numerical Example

(Parameters based on "Expert Opinion")

θ	X=1 (Poor)	X=1 (Okay)	X=1 (Good)
Low	.70	.25	.05
Med	.35	.40	.25
High	.10	.40	.50

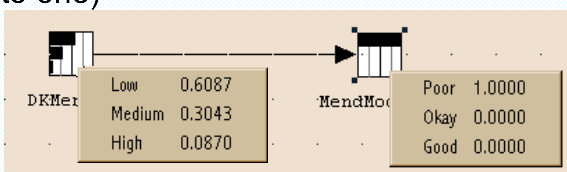
Likelihood function for possible values of θ, when it is not known but X=Poor has been observed. (Note: they don't add up to one)

A Simple Numerical Example

(Parameters based on "Expert Opinion")

θ	X=1 (Poor)	X=1 (Okay)	X=1 (Good)
Low	.70	.25	.05
Med	.35	.40	.25
High	.10	.40	.50

Likelihood function for possible values of θ, when it is not known but X=Poor has been observed. (Note: they don't add up to one)

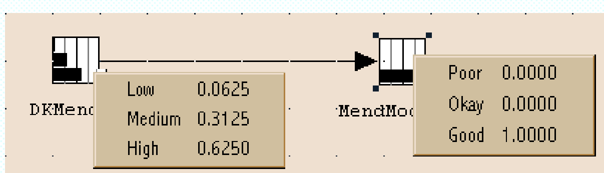


A Simple Numerical Example

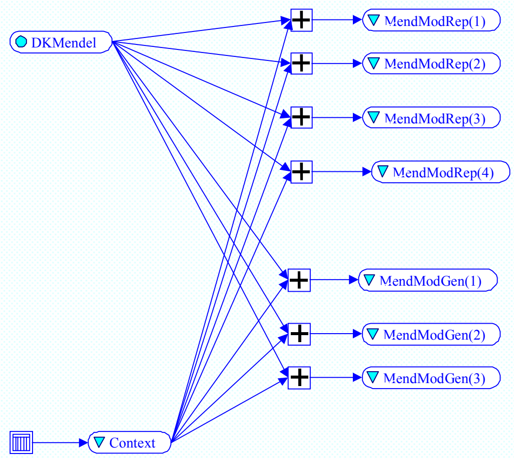
(Parameters based on "Expert Opinion")

θ	X=1 (Poor)	X=1 (Okay)	X=1 (Good)
Low	.70	.25	.05
Med	.35	.40	.25
High	.10	.40	.50

Likelihood function for possible values of θ, when it is not known but X=Good has been observed.

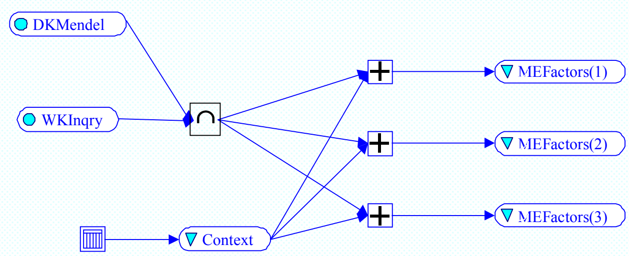


Measurement-Model fragments in EMs used in Agouti 1



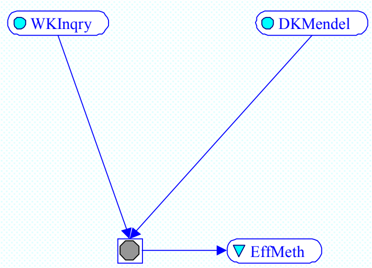
DKMendel is the student-model variable that determines probabilities of response to the several observable variables in the Mode of Inheritance chart. "Context" is a parent that induces conditional dependence among these observations, above relationships due to DKMendel.

Measurement-Model fragments in EMs used in Agouti 1



WKInqry and DKMendel are required *conjunctively* to determine probabilities for the observable variables concerning what can be inferred about mode of inheritance from field populations, offspring from crossing them, and second-generation crosses. Again the observables are modeled as conditionally dependent.

Measurement-Model fragments in EMs used in Agouti 1



DKMendel is an *inhibitor* of WKInqry for determine probabilities for the observable variable "what to do next?" Once a student is at least at the 'medium' level of DKMendel she knows enough to meaningfully engage the issue; then probabilities increase with WKInqry.

Expression in Equations

$\Pr(S_i)$ is an SM-BIN fragment.

$\Pr(X_{im} | S_{im}^{(m)}, \pi_m)$ is an EM-BIN fragment for Task *m* --

π_m represents conditional probabilities of observable response variables X_{im} given SM 'footprint' $S_{im}^{(m)}$.

S_i and X_{im}'s may be multivariate, as they are in Agouti 1.

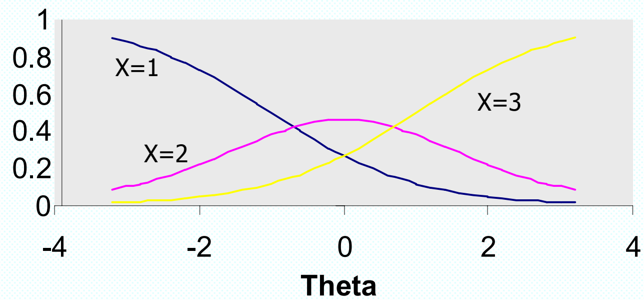
Bayes net combining SM and EM fragments for *M* tasks:

$$\Pr\left(S_i, X_{i1}, \ldots, X_{iM} \middle| \pi_1 .. \pi_M\right) = \Pr(S_i) \prod_{m=1}^{M} \Pr\left(X_{im} \middle| S_i^{(m)}, \pi_M\right)$$

Conditional Probabilities

- Structure by design
  - » Compensatory, Conjunctive, Disjunctive, Inhibitor
  - » Conditional dependence in complex performances
- Initial values from experts & task features
- Refined with data (MCMC estimation)

The "Effective θ" Method (1): Samejima's Model



a_j=1
b_{j1}=-1
b_{j2}=+1

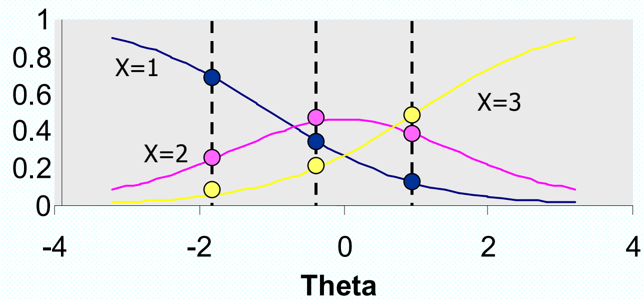
Samejima's (1969) psychometric model for graded responses:

$$\Pr(X_{ij} \geq k|\theta) = \text{logit}^{-1}(a_j(\theta + b_{jk}));$$

$$\Pr(X_{ij} = k|\theta) = \Pr(X_{ij} \geq k|\theta) - \Pr(X_{ij} \geq k+1|\theta).$$

The "Effective θ" Method (2):

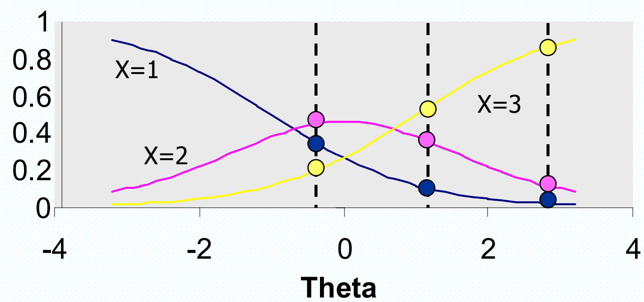
Conditional Probabilities for Three θ's



θ	X=1 (Poor)	X=1 (Okay)	X=1 (Good)
Low= -1.8	.70	.25	.05
Med= -.4	.35	.40	.25
High= 1.0	.10	.40	.50

The "Effective θ" Method (2):

Conditional Probabilities for Three Different θ's



θ	X=1 (Poor)	X=1 (Okay)	X=1 (Good)
Low= -.2	.35	.40	.25
Med= .6	.10	.35	.55
High= 1.4	.05	.10	.90

Parameterization for Compensatory Relationship

Initial conditional probability distributions for Observable 1 of Task 1

DKMendel index[1]	θ_{11}[2]	Context index[3]	θ_{11}^*[4]	Pr(X=k) Low	Medium	High
-1	0.00	-1	-0.50	0.50	0.23	0.27
-1	0.00	1	0.50	0.27	0.23	0.50
0	1.00	-1	0.50	0.27	0.23	0.50
0	1.00	1	1.50	0.12	0.15	0.73
1	2.00	-1	1.50	0.12	0.15	0.73
1	2.00	1	2.50	0.05	0.07	0.88

[1] Low=-1, Medium=0, High=1

[2] $\theta_{11} = c_{11} i_{DKM} + d_{11} = 1.00\, i_{DKM} + 1.00$

[3] Low=-1, High=1

[4] $\theta_{11}^* = \theta_{11} + e_{11} i_{C_{Ed1}} = \theta_{11} + .5\, i_{EM}$

Map functions of indices for levels of SM parents onto a single effective θ scale--propensity to respond well on this observable.
Here, two linear functions.

Parameterization for Compensatory Relationship

Initial conditional probability distributions for Observable 1 of Task 1

DKMendel index[1]	θ_{11}[2]	Context index[3]	θ_{11}^*[4]	Pr(X=k) Low	Medium	High
-1	0.00	-1	-0.50	0.50	0.23	0.27
-1	0.00	1	0.50	0.27	0.23	0.50
0	1.00	-1	0.50	0.27	0.23	0.50
0	1.00	1	1.50	0.12	0.15	0.73
1	2.00	-1	1.50	0.12	0.15	0.73
1	2.00	1	2.50	0.05	0.07	0.88

[1] Low=-1, Medium=0, High=1

[2] $\theta_{11} = c_{11} i_{DKM} + d_{11} = 1.00\, i_{DKM} + 1.00$

[3] Low=-1, High=1

[4] $\theta_{11}^* = \theta_{11} + e_{11} i_{C_{Ed1}} = \theta_{11} + .5\, i_{EM}$

○ Parameters initially set by expert opinion (or TM values); to be refined with info from pretest data. Experts expected this one to be easy.

Updating parameters with MCMC estimation

- The full Bayesian model
- MCMC estimation (Gibbs sampling)
- Our data set
- A word about testing fit

The Full Bayesian Model

$$p(\mathbf{X},\mathbf{S},\text{ð},\text{ç},\lambda)=\prod_i \prod_m \prod_j p\left(x_{imj} \middle| s_i^{(m)},\pi_{mj}\right) p\left(\pi_{mj} \middle| \eta_m\right) p(\eta_m)\, p\left(s_i \middle| \lambda\right) p(\lambda).$$

- λ: hyperparameters for SM var distributions
- ç: hyperparameter for conditional probability distributions in the effective θ models
- ð: conditional probs, all observables *j* in tasks *m*
- $\mathbf{S}$: SM variables for all students *i*
- $\mathbf{X}$: observable responses *j* in *m* for all students *i*

The Full Bayesian Model

$$p(\mathbf{X},\mathbf{S},\text{ð},\text{ç},\lambda)=\prod_i\prod_m\prod_j \underbrace{p\left(x_{imj}\middle|s_i^{(m)},\pi_{mj}\right)}\underbrace{p\left(\pi_{mj}\middle|\eta_m\right)}\underbrace{p(\eta_m)}\underbrace{p(s_i|\lambda)}\underbrace{p(\lambda)}.$$

Conditional distributions of observables, given SM vars & conditional probs

Distributions of conditional probs, given their hyperparameters

Dists. of hyperparameters for conditional probs

Distributions of students' SM var values, given their hyperparameters

Distributions of hyperparameters for SM variables

MCMC estimation (Gibbs sampling)

- The full Bayesian distribution is a joint distribution of *all data* and *all the parameters at all levels*.
- If you knew the values of any subset of data & parameters, you could (in principle) calculate the conditional distribution of the remaining ones.
- In each iteration of a Gibbs sampler, for a given parameter, you hold draws for all the other parameters and data fixed, and draw a value from the (full) conditional of that parameter.
- Broadly, the distributions of these draws are the same as the posterior distributions for the parameters.

MCMC estimation (Gibbs sampling)

We're interested in $p(\mathbf{S},\pi,\eta,\lambda|\mathbf{X})$.

The $t+1^{\text{th}}$ iteration looks like this:

- Draw $\mathbf{S}^{t+1}$ from $p(\mathbf{S}|\pi^t,\eta^t,\lambda^t,\mathbf{X})$;
- Draw η^{t+1} from $p(\eta|\mathbf{S}^{t+1},\pi^t,\lambda^t,\mathbf{X})$;
- Calculate π^{t+1} from η^{t+1}; and
- Draw λ^{t+1} from $p(\lambda|\mathbf{S}^{t+1},\pi^{t+1},\eta^{t+1},\mathbf{X})$.

Spiegelhalter, Thomas, Best, & Gilks (1995)*: BUGS*.
http://www.mrc-bsu.cam.ac.uk/bugs/

Our data set

- 28 summer students worked through Agouti 1.
- Disadvantage in that the task "dropped in out of the sky" to them; weren't familiar with expectations, interface, knowledge representations.
- Students ranged from a few with mostly Good scores to some students with all Poor.
- Observations ranged from mostly Good to all Poor.
- Not many Okay values for observables.

An Updated Conditional Probability Matrix

Revised conditional probability table for Observable 4 of Task 1

DKMendel index[1]	θ_{11}[2]	*Context* index[3]	θ_{11}^{*}[4]	Pr(X=k) Low	Medium	High
-1	0.00	-1	-0.50	0.98	0.01	0.01
-1	0.00	1	0.50	0.93	0.05	0.03
0	1.00	-1	0.50	0.96	0.03	0.02
0	1.00	1	1.50	0.85	0.09	0.06
1	2.00	-1	1.50	0.91	0.06	0.04
1	2.00	1	2.50	0.72	0.16	0.13

[1] Low=-1, Medium=0, High=1

[2] $\theta_{11} = c_{11} i_{DKM} + d_{11} = 0.80\, i_{DKM} + -2.92$

(1.29) (0.00)

[3] Low=-1, High=1

[4] $\theta_{11}^{*} = \theta_{11} + e_{11} i_{C_{Ew1}} = \theta_{11} + 0.69\, i_{EM}$

(1.29)

Testing Fit

- We don't expect to *learn* the structure of the models and parameter values from data cold.
  - > Depend on correspondence between task design and statistical structure
  - > Take advantage of expert opinion for structure and for initial distributions.
- MCMC fit strategy (Rubin, Gelman)
  - > Generate "shadow" data, modeled to have same distribution as that posited for actual data
  - > Essentially a tailored null distribution for any statistics
  - > Compare key features of real data to those of shadow data

Future work

- This application,
  > More students
  > More segments
  > Compare structures: e.g., conjunctive vs compensatory
- Generally,
  > Conditional probabilities functions of TM variables
  > Automatic generation of BUGS code
  > Approximations for handling large cliques