Hierarchical Models for Student Grading

Joachim Vandekerckhove and Michael Lee

Student Grading

- Five students have completed all or some of the 50 questions used to determine their course grade

Student	Correct	Completed	Percentage
One	39	50	78%
Two	47	50	94%
Three	40	50	80%
Four	8	10	80%
Five	5	10	50%
Six	–	–	–

Student Grading

- Five students have completed all or some of the 50 questions used to determine their course grade
  - A sixth student has not yet started the questions

Student	Correct	Completed	Percentage
One	39	50	78%
Two	47	50	94%
Three	40	50	80%
Four	8	10	80%
Five	5	10	50%
Six	–	–	–

Student Grading

- Five students have completed all or some of the 50 questions used to determine their course grade
  - A sixth student has not yet started the questions

Student	Correct	Completed	Percentage
One	39	50	78%
Two	47	50	94%
Three	40	50	80%
Four	8	10	80%
Five	5	10	50%
Six	–	–	–

- We are interested in the underlying ability of students to answer questions correctly, and predictions about their final grade

Independent Rate Model

- One reasonable model is a rate model that assumes the y_i correct answers out of n_i total questions for the ith student are generated by an underlying probability θ_i, so that

$$y_i \sim \text{binomial}(\theta_i, n_i)$$

$$\begin{tikzcd}
\theta_1 \arrow[d] & \theta_2 \arrow[d] & \theta_3 \arrow[d] & & \theta_n \arrow[d] \\
y_1 & y_2 & y_3 & & y_n
\end{tikzcd}$$

Independent Rate Model

- One reasonable model is a rate model that assumes the y_i correct answers out of n_i total questions for the ith student are generated by an underlying probability θ_i, so that

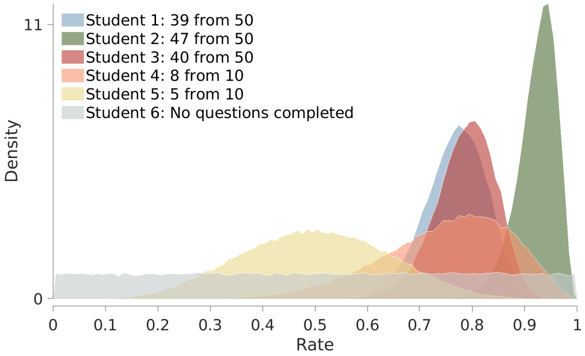
$$y_i \sim \text{binomial}(\theta_i, n_i)$$

- The underlying rates are independent of each other, and given the uniform prior $\theta_i \sim \text{uniform}(0, 1)$

$$\begin{tikzcd} \theta_1 \arrow[d] & \theta_2 \arrow[d] & \theta_3 \arrow[d] & \theta_n \arrow[d] \\ y_1 & y_2 & y_3 & y_n \end{tikzcd}$$

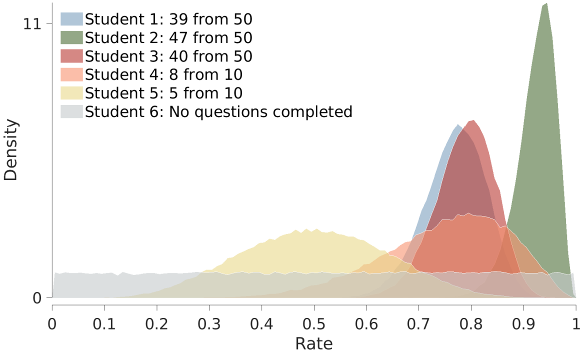
Independent Rate Model Inferences

- The posterior distributions for the θ_i are shown



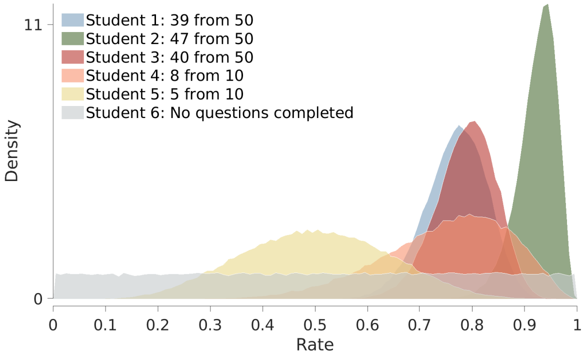
Independent Rate Model Inferences

- The posterior distributions for the θ_i are shown
  - Certainty depends on the number of questions completed



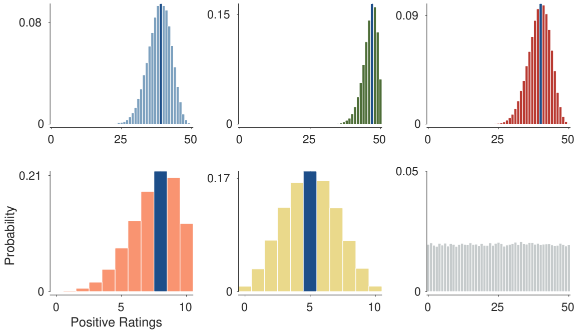
Independent Rate Model Inferences

- The posterior distributions for the θ_i are shown
  - Certainty depends on the number of questions completed
  - The posterior distribution for 6 is simply the prior distribution



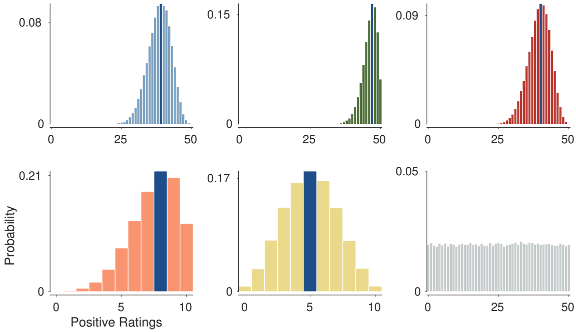
Independent Rate Model Posterior Predictions

- A posterior predictive analysis shows the model describes the observed data well



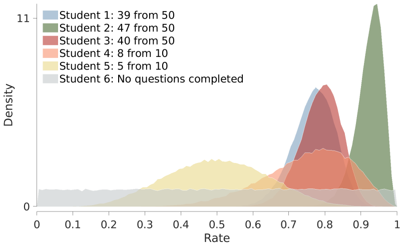
Independent Rate Model Posterior Predictions

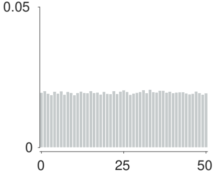
- A posterior predictive analysis shows the model describes the observed data well
  - The prediction for 6 is really a prior prediction, and expects all final results between 0 and 50 out of 50 to be equally likely



Intuitions

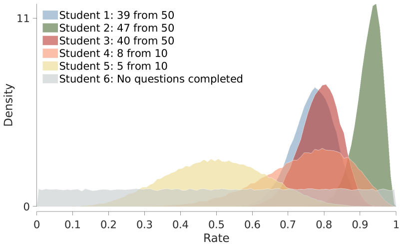
- Do the posterior inferences for 5 and 6 seem the most reasonable? Does the prediction for 6 seem the most reasonable?

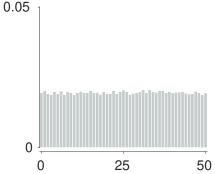




Intuitions

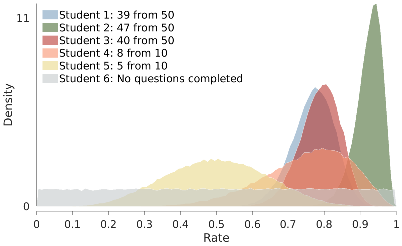
- Do the posterior inferences for 5 and 6 seem the most reasonable? Does the prediction for 6 seem the most reasonable?
    - It seems conceivable, even likely, that 5 just got off to a bad start, and will end up having a grade above 50% correct

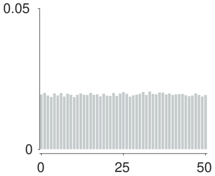




Intuitions

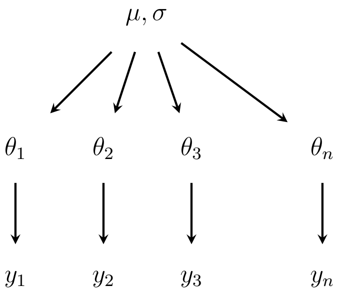
- Do the posterior inferences for 5 and 6 seem the most reasonable? Does the prediction for 6 seem the most reasonable?
  - It seems conceivable, even likely, that 5 just got off to a bad start, and will end up having a grade above 50% correct
  - Even though 6 has not answered any questions, it seems likely, for example, their accuracy rate will be above 0.5 rather than below 0.5, and they will likely also score better than 50%





Hierarchical Model

- The mismatch between intuition and the model predictions arise because of the assumption students are independent



Hierarchical Model

- The mismatch between intuition and the model predictions arise because of the assumption students are independent
- Hierarchical models allow both sameness and difference to be modeled, using individual-level parameters that are connected by all being drawn from an over-arching group distribution

$$
\begin{tikzcd}
 & \mu, \sigma \arrow[dl] \arrow[d] \arrow[dr] \arrow[drr] & & \\
\theta_1 \arrow[d] & \theta_2 \arrow[d] & \theta_3 \arrow[d] & \theta_n \arrow[d] \\
y_1 & y_2 & y_3 & y_n
\end{tikzcd}
$$

Hierarchical Model

- The mismatch between intuition and the model predictions arise because of the assumption students are independent
- Hierarchical models allow both sameness and difference to be modeled, using individual-level parameters that are connected by all being drawn from an over-arching group distribution
  - For this example, the hierarchical distribution is the class curve

$$
\begin{tikzcd}
 & & \mu, \sigma \arrow[dll] \arrow[dl] \arrow[d] \arrow[dr] & \\
\theta_1 \arrow[d] & \theta_2 \arrow[d] & \theta_3 \arrow[d] & \theta_n \arrow[d] \\
y_1 & y_2 & y_3 & y_n
\end{tikzcd}
$$

Hierarchical Rate Model

- Assume all of the individual student rates θ_i come from a (truncated) Gaussian group distribution with mean (mode) μ and standard deviation σ

$$\theta_i \sim \mathrm{Gaussian}(\mu, \sigma^2)\mathrm{T}\,(0, 1)$$

Hierarchical Rate Model

- Assume all of the individual student rates θ_i come from a (truncated) Gaussian group distribution with mean (mode) μ and standard deviation σ

$$\theta_i \sim \mathrm{Gaussian}(\mu, \sigma^2)\mathrm{T}\,(0,1)$$

- The hyper-parameters are given priors $\mu \sim \mathrm{uniform}(0,1)$ and $\sigma \sim \mathrm{uniform}(0,1)$

Hierarchical Rate Model

- Assume all of the individual student rates θ_i come from a (truncated) Gaussian group distribution with mean (mode) μ and standard deviation σ

$$\theta_i \sim \mathrm{Gaussian}(\mu, \sigma^2)\mathrm{T}\,(0, 1)$$

- The hyper-parameters are given priors $\mu \sim \mathrm{uniform}(0, 1)$ and $\sigma \sim \mathrm{uniform}(0, 1)$
- The students are now hierarchically related to each other, through their shared membership of group distribution

Hierarchical Rate Model

- Hierarchical models allow inferences at multiple levels of abstraction

Hierarchical Rate Model

- Hierarchical models allow inferences at multiple levels of abstraction
    - Inferences about μ and σ are inferences about the class curve

Hierarchical Rate Model

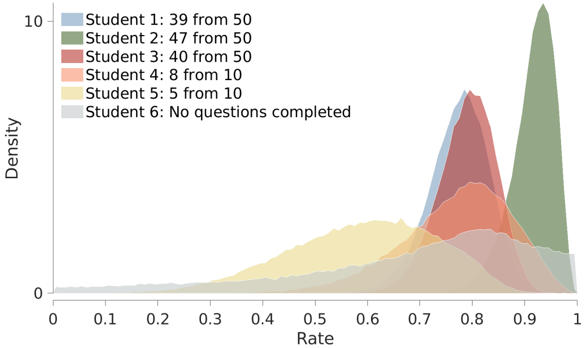
- Hierarchical models allow inferences at multiple levels of abstraction
  - Inferences about μ and σ are inferences about the class curve
  - Inferences about θ_i are inferences about individual students

Hierarchical Rate Model

- Hierarchical models allow inferences at multiple levels of abstraction
    - Inferences about μ and σ are inferences about the class curve
    - Inferences about θ_i are inferences about individual students
- Hierarchical models of individual differences can capture both what people have in common, and their differences

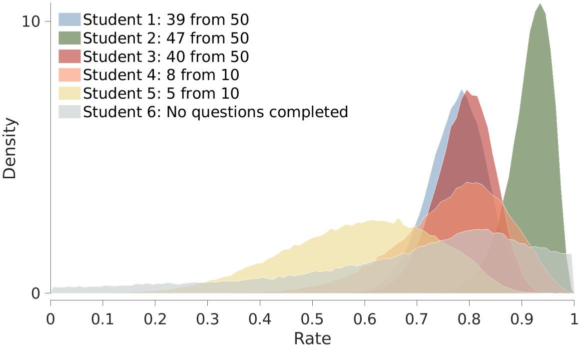
Hierarchical Rate Model Individual Inferences

- The inferences for θ_i based on the hierarchical model are shown



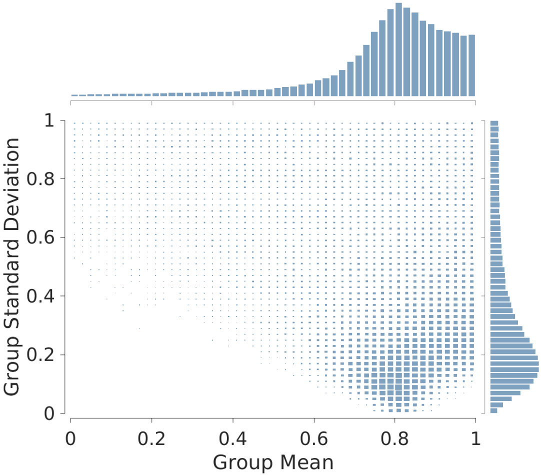
Hierarchical Rate Model Individual Inferences

- The inferences for θ_i based on the hierarchical model are shown
  - The posterior distributions for all students, but noticeably students five and six, are influenced by the group distribution by an effect called "shrinkage" (or "sharing statistical strength")



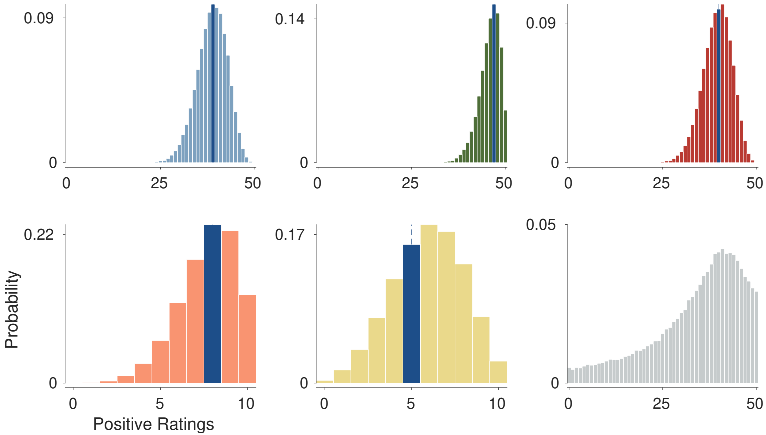
Hierarchical Rate Group Inferences

- The joint and marginal posterior distributions for the group-level μ and σ parameters are shown



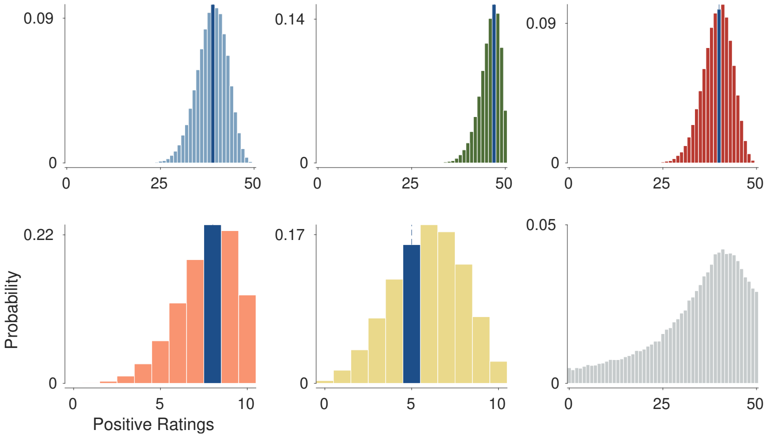
Hierarchical Rate Model Posterior Predictions

- The posterior predictive analysis shows the hierarchical model continues to describe the observed data well

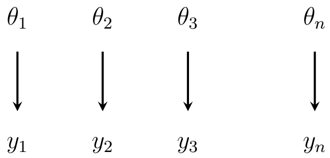


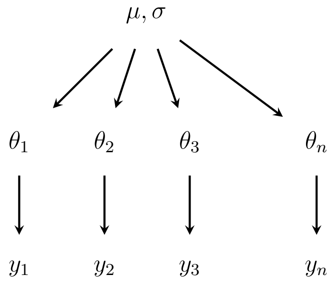
Hierarchical Rate Model Posterior Predictions

- The posterior predictive analysis shows the hierarchical model continues to describe the observed data well
  - The prediction for 6 is now based on the group distribution



Key Points





- Hierarchical models extend a basic cognitive model to include an account of how the parameters themselves are generated

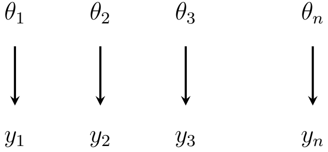
Key Points

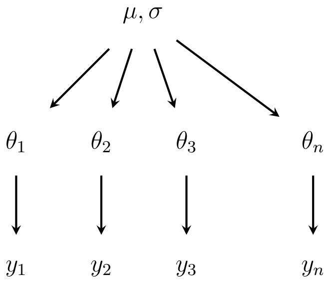
$$
\begin{tikzcd}
\theta_1 \arrow[d] & \theta_2 \arrow[d] & \theta_3 \arrow[d] & \theta_n \arrow[d] \\
y_1 & y_2 & y_3 & y_n
\end{tikzcd}
$$

$$
\begin{tikzcd}
& & \mu, \sigma \arrow[dll] \arrow[dl] \arrow[dr] \arrow[drr] & & \\
\theta_1 \arrow[d] & \theta_2 \arrow[d] & & \theta_3 \arrow[d] & \theta_n \arrow[d] \\
y_1 & y_2 & & y_3 & y_n
\end{tikzcd}
$$

- Hierarchical models extend a basic cognitive model to include an account of how the parameters themselves are generated
- One common application is to individual differences, allowing inference at both the individual and group level

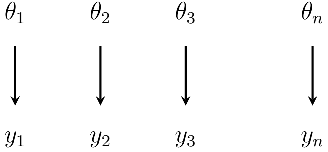
Key Points

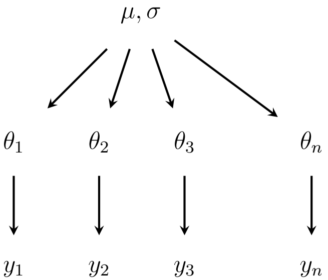




- Hierarchical models extend a basic cognitive model to include an account of how the parameters themselves are generated
- One common application is to individual differences, allowing inference at both the individual and group level
- Hierarchical models make different inferences and predictions, because they make different assumptions

Key Points





- Hierarchical models extend a basic cognitive model to include an account of how the parameters themselves are generated
- One common application is to individual differences, allowing inference at both the individual and group level
- Hierarchical models make different inferences and predictions, because they make different assumptions
  - whether the hierarchical model inferences are better or worse depends on the usefulness of the assumptions