Bayesian IRT for the Masses

S. McKay Curtis

University of Washington
Department of Statistics

August 30, 2010

Acknowledgments

Disclaimers?

- Funded in part by Grant R13AG030995-01A1 from the National Institute on Aging
- The views expressed in written conference materials or publications and by speakers and moderators do not necessarily reflect the official policies of the Department of Health and Human Services; nor does mention by trade names, commercial practices, or organizations imply endorsement by the U.S. Government.

Outline

Outline

1. Bayesian Inference
2. Item Response Theory
3. Bayesian Item Response Theory
4. Longitudinal Bayesian Item Response Theory

What is Probability?

What is Probability?

Consider the following statements...

What is Probability?

Consider the following statements. . .

1. The probability of a coin toss landing heads is 1/2. John Kerrich flipped a coin 10,000 times while a POW in WWII. He obtained 5,067 heads.

What is Probability?

Consider the following statements. . .

1. The probability of a coin toss landing heads is 1/2. John Kerrich flipped a coin 10,000 times while a POW in WWII. He obtained 5,067 heads.
2. The probability that the Republicans will control the House of Representatives after the 2010 Congressional elections is 0.735. Intrade odds, 26 Aug 2010, 10:50am

What is Probability?

Consider the following statements...

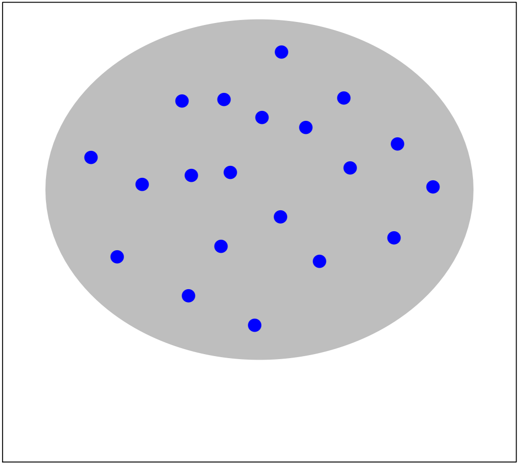
1. The probability of a coin toss landing heads is 1/2. John Kerrich flipped a coin 10,000 times while a POW in WWII. He obtained 5,067 heads.
2. The probability that the Republicans will control the House of Representatives after the 2010 Congressional elections is 0.735. Intrade odds, 26 Aug 2010, 10:50am
3. The probability that James Madison wrote the disputed Federalist papers is > 0.999. Mosteller and Wallace (1964).

What is Probability?

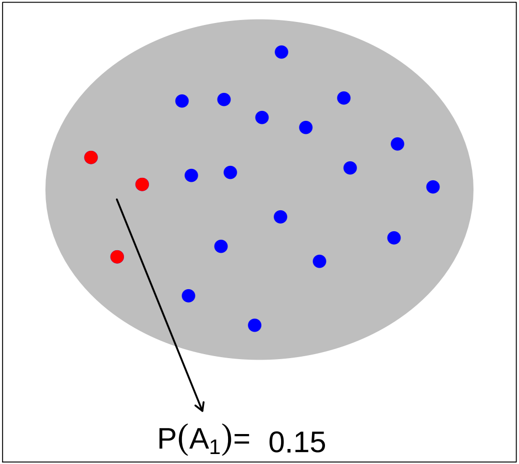
Consider the following statements...

1. The probability of a coin toss landing heads is 1/2. John Kerrich flipped a coin 10,000 times while a POW in WWII. He obtained 5,067 heads.
2. The probability that the Republicans will control the House of Representatives after the 2010 Congressional elections is 0.735. Intrade odds, 26 Aug 2010, 10:50am
3. The probability that James Madison wrote the disputed Federalist papers is > 0.999. Mosteller and Wallace (1964).
4. The probability that God exists is 0.67. Stephen D. Unwin, The Probability of God: A Simple Calculation that Proves the Ultimate Truth.

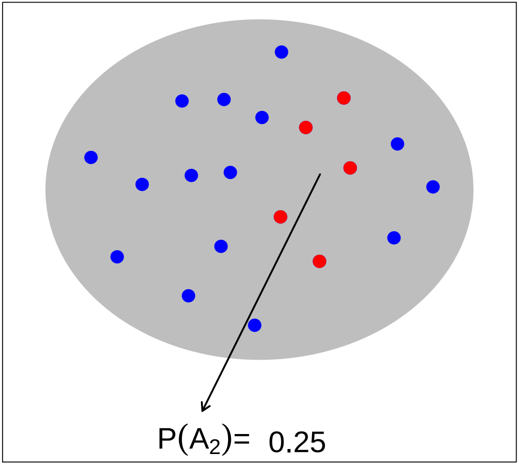
What is Probability?



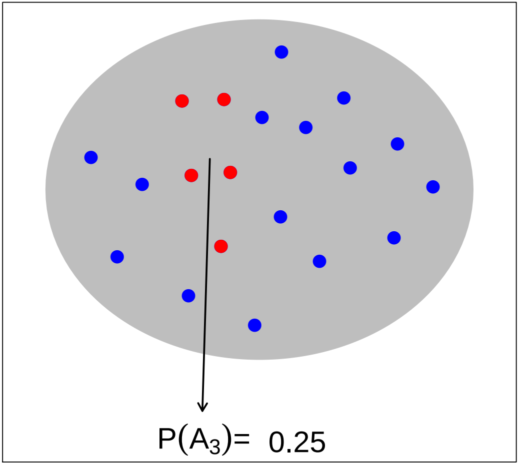
What is Probability?



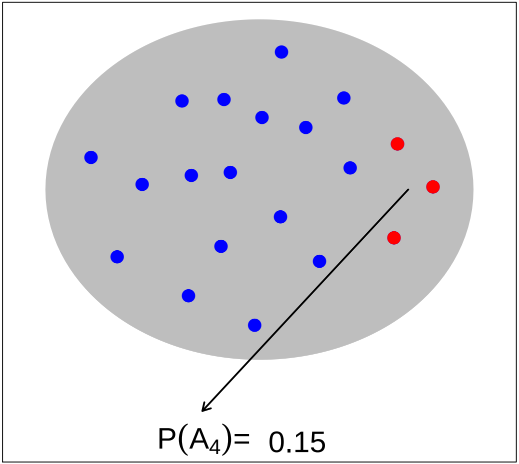
What is Probability?



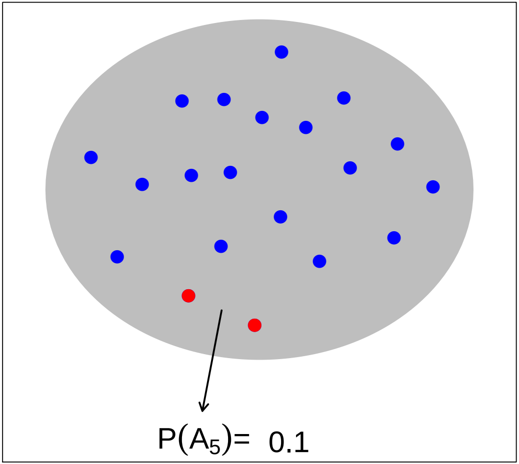
What is Probability?



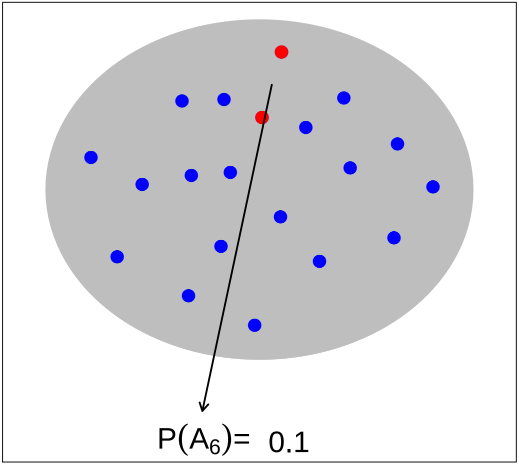
What is Probability?



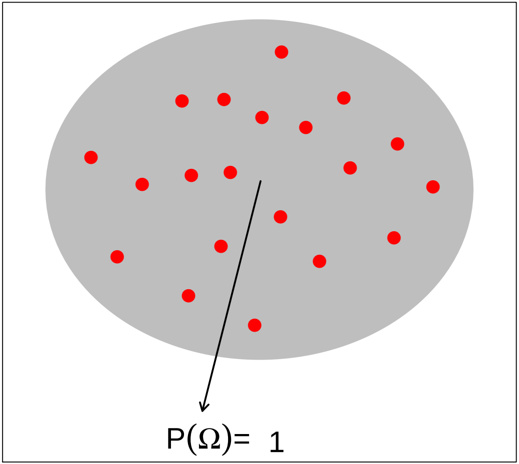
What is Probability?



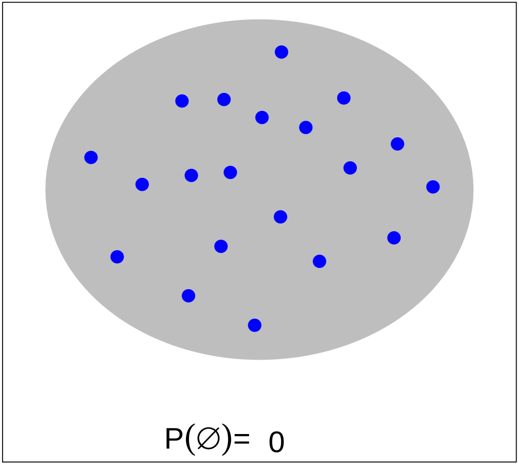
What is Probability?



What is Probability?



What is Probability?



What is probability?

Technical definition

What is probability?

Technical definition

Definition: Probability

Probability is a set function $P(\cdot)$ defined on subsets of a space Ω that satisfies the following properties:

1. $P(\Omega) = 1$
2. For a subset $A \subset \Omega$, $P(A) \geq 0$
3. If $A_1, A_2, \ldots$ are disjoint subsets of Ω then

$$P(A_1 \cup A_2 \cup \cdots) = P(A_1) + P(A_2) + \cdots$$

What should we use probability for?

Bayesian statisticians vs. Frequentist statisticians

If probability is just a set function with special properties, then the question is “What should we use probability for?”

What should we use probability for?

Bayesian statisticians vs. Frequentist statisticians

If probability is just a set function with special properties, then the question is "What should we use probability for?"

- A frequentist statistician[1]...

What should we use probability for?

Bayesian statisticians vs. Frequentist statisticians

If probability is just a set function with special properties, then the question is "What should we use probability for?"

- A frequentist statistician[1]...

[1] Obviously, this is just a caricature.

What should we use probability for?

Bayesian statisticians vs. Frequentist statisticians

If probability is just a set function with special properties, then the question is "What should we use probability for?"

- A frequentist statistician[1]...
  - will use probability only to model uncertainty in the outcomes of repeatable experiments (e.g., like the toss of a coin).

[1] Obviously, this is just a caricature.

What should we use probability for?

Bayesian statisticians vs. Frequentist statisticians

If probability is just a set function with special properties, then the question is "What should we use probability for?"

- A frequentist statistician[1]...
  - will use probability only to model uncertainty in the outcomes of repeatable experiments (e.g., like the toss of a coin).
  - believes probability is an objective property—the long-run relative frequency (hence the name "frequentist")—of some process that generates data.

[1]Obviously, this is just a caricature.

What should we use probability for?

Bayesian statisticians vs. Frequentist statisticians

If probability is just a set function with special properties, then the question is "What should we use probability for?"

- A frequentist statistician[1]...
  - will use probability only to model uncertainty in the outcomes of repeatable experiments (e.g., like the toss of a coin).
  - believes probability is an objective property—the long-run relative frequency (hence the name "frequentist")—of some process that generates data.
- A Bayesian statistician[2]...

[1] Obviously, this is just a caricature.

What should we use probability for?

Bayesian statisticians vs. Frequentist statisticians

If probability is just a set function with special properties, then the question is "What should we use probability for?"

- A frequentist statistician[1]...
  - will use probability only to model uncertainty in the outcomes of repeatable experiments (e.g., like the toss of a coin).
  - believes probability is an objective property—the long-run relative frequency (hence the name "frequentist")—of some process that generates data.
- A Bayesian statistician[2]...

[1]Obviously, this is just a caricature.

[2]And so is this.

What should we use probability for?

Bayesian statisticians vs. Frequentist statisticians

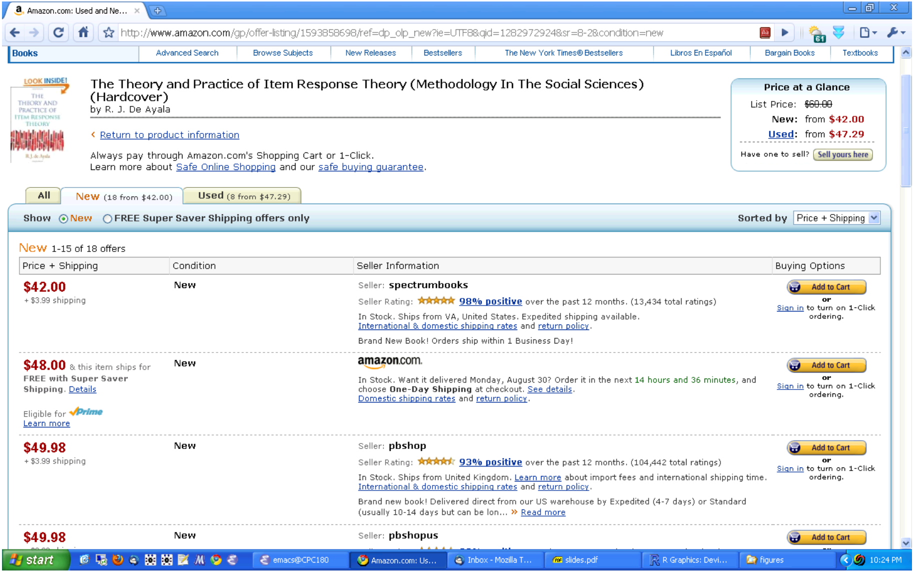
If probability is just a set function with special properties, then the question is "What should we use probability for?"

- A frequentist statistician[1]. . .
  - will use probability only to model uncertainty in the outcomes of repeatable experiments (e.g., like the toss of a coin).
  - believes probability is an objective property—the long-run relative frequency (hence the name "frequentist")—of some process that generates data.
- A Bayesian statistician[2]. . .
  - will use probability to model uncertainty from any source (e.g., that a coin toss lands heads or that Madison wrote the disputed Federalist papers.)

[1]Obviously, this is just a caricature.
[2]And so is this.

Bayesian vs. Frequentist inference in practice

An example using amazon.com seller ratings



Bayesian vs. Frequentist inference in practice

An example using amazon.com seller ratings

- You want to buy a used book on amazon.com

Bayesian vs. Frequentist inference in practice

An example using amazon.com seller ratings

- You want to buy a used book on amazon.com
- Your book is being sold by two "marketplace" sellers for the same price and in the same condition.

Bayesian vs. Frequentist inference in practice

An example using amazon.com seller ratings

- You want to buy a used book on amazon.com
- Your book is being sold by two "marketplace" sellers for the same price and in the same condition.

Marketplace sellers

Name	Positive Reviews	Total Reviews	Percent Positive
BigNBooks	29	30	96.7%
LittleNBooks	5	5	100%

Bayesian vs. Frequentist inference in practice

An example using amazon.com seller ratings

- You want to buy a used book on amazon.com
- Your book is being sold by two “marketplace” sellers for the same price and in the same condition.

Marketplace sellers

Name	Positive Reviews	Total Reviews	Percent Positive
BigNBooks	29	30	96.7%
LittleNBooks	5	5	100%

- Which book seller should you choose?

Bayesian vs. Frequentist inference in practice

An example using amazon.com seller ratings

- You want to buy a used book on amazon.com
- Your book is being sold by two “marketplace” sellers for the same price and in the same condition.

Marketplace sellers

Name	Positive Reviews	Total Reviews	Percent Positive
BigNBooks	29	30	96.7%
LittleNBooks	5	5	100%

- Which book seller should you choose?
  - LittleNBooks has a higher rating, but. . .

Bayesian vs. Frequentist inference in practice

An example using amazon.com seller ratings

- You want to buy a used book on amazon.com
- Your book is being sold by two "marketplace" sellers for the same price and in the same condition.

Marketplace sellers

Name	Positive Reviews	Total Reviews	Percent Positive
BigNBooks	29	30	96.7%
LittleNBooks	5	5	100%

- Which book seller should you choose?
  - LittleNBooks has a higher rating, but...
  - LittleNBooks only has five ratings.

Bayesian vs. Frequentist inference in practice

A mathematical model for bookseller ratings

- Reviewers' ratings = repeatable event

Bayesian vs. Frequentist inference in practice

A mathematical model for bookseller ratings

- Reviewers' ratings = repeatable event
    - Frequentists ♡ probability.

Bayesian vs. Frequentist inference in practice

A mathematical model for bookseller ratings

- Reviewers' ratings = repeatable event
  - Frequentists ♡ probability.
  - Bayesians ♡ probability.

Bayesian vs. Frequentist inference in practice

A mathematical model for bookseller ratings

- Reviewers' ratings = repeatable event
  - Frequentists ♡ probability.
  - Bayesians ♡ probability.
- A simple model posits that there exists some true probability of a positive review for BigNBooks (θ_B) and LittleNBooks (θ_L).

Bayesian vs. Frequentist inference in practice

A mathematical model for bookseller ratings

- Reviewers' ratings = repeatable event
  - Frequentists ♡ probability.
  - Bayesians ♡ probability.
- A simple model posits that there exists some true probability of a positive review for BigNBooks (θ_B) and LittleNBooks (θ_L).
  - $Y_i^{(B)} = \begin{cases} 1 & \text{if review positive} \\ 0 & \text{if review negative} \end{cases}$

Bayesian vs. Frequentist inference in practice

A mathematical model for bookseller ratings

- Reviewers' ratings = repeatable event
  - Frequentists ♡ probability.
  - Bayesians ♡ probability.
- A simple model posits that there exists some true probability of a positive review for BigNBooks (θ_B) and LittleNBooks (θ_L).
  - $Y_i^{(B)} = \begin{cases} 1 & \text{if review positive} \\ 0 & \text{if review negative} \end{cases}$
  - $P(Y_i^{(B)} = 1|\theta_B) = \theta_B$

Bayesian vs. Frequentist inference in practice

A mathematical model for bookseller ratings

- Reviewers' ratings = repeatable event
  - Frequentists ♡ probability.
  - Bayesians ♡ probability.
- A simple model posits that there exists some true probability of a positive review for BigNBooks (θ_B) and LittleNBooks (θ_L).
  - $Y_i^{(B)} = \begin{cases} 1 & \text{if review positive} \\ 0 & \text{if review negative} \end{cases}$
  - $P(Y_i^{(B)} = 1|\theta_B) = \theta_B$
  - $Y_i^{(L)} = \begin{cases} 1 & \text{if review positive} \\ 0 & \text{if review negative} \end{cases}$

Bayesian vs. Frequentist inference in practice

A mathematical model for bookseller ratings

- Reviewers' ratings = repeatable event
  - Frequentists ♡ probability.
  - Bayesians ♡ probability.
- A simple model posits that there exists some true probability of a positive review for BigNBooks (θ_B) and LittleNBooks (θ_L).
  - $Y_i^{(B)} = \begin{cases} 1 & \text{if review positive} \\ 0 & \text{if review negative} \end{cases}$
  - $P(Y_i^{(B)} = 1|\theta_B) = \theta_B$
  - $Y_i^{(L)} = \begin{cases} 1 & \text{if review positive} \\ 0 & \text{if review negative} \end{cases}$
  - $P(Y_i^{(L)} = 1|\theta_L) = \theta_L$

Bayesian vs. Frequentist inference in practice

The likelihood

- Assume $Y_1^{(B)}, \ldots, Y_{30}^{(B)} = \mathbf{Y}^{(B)}$ and $Y_1^{(L)}, \ldots, Y_5^{(L)} = \mathbf{Y}^{(L)}$ independent, then

Bayesian vs. Frequentist inference in practice

The likelihood

- Assume $Y_1^{(B)}, \ldots, Y_{30}^{(B)} = \mathbf{Y}^{(B)}$ and $Y_1^{(L)}, \ldots, Y_5^{(L)} = \mathbf{Y}^{(L)}$ independent, then
-

$$\begin{aligned} P(Y_1^{(B)} = 1, \ldots, Y_{29}^{(B)} = 1, Y_{30}^{(B)} = 0 | \theta_B) &= P(\mathbf{Y}^{(B)} | \theta_B) \\ &= \theta_B^{29}(1 - \theta_B) \end{aligned}$$

Bayesian vs. Frequentist inference in practice

The likelihood

- Assume $Y_1^{(B)}, \ldots, Y_{30}^{(B)} = \mathbf{Y}^{(B)}$ and $Y_1^{(L)}, \ldots, Y_5^{(L)} = \mathbf{Y}^{(L)}$ independent, then
- $$\begin{aligned} P(Y_1^{(B)} = 1, \ldots, Y_{29}^{(B)} = 1, Y_{30}^{(B)} = 0 | \theta_B) &= P(\mathbf{Y}^{(B)} | \theta_B) \\ &= \theta_B^{29}(1 - \theta_B) \end{aligned}$$
- $$\begin{aligned} P(Y_1^{(L)} = 1, \ldots, Y_5^{(L)} = 1 | \theta_L) &= P(\mathbf{Y}^{(L)} | \theta_L) \\ &= \theta_L^5 \end{aligned}$$

Bayesian vs. Frequentist inference in practice

Frequentist inference

- θ_B and θ_L are fixed, unknown constants $\rightarrow$ ~~probability~~

Bayesian vs. Frequentist inference in practice

Frequentist inference

- θ_B and θ_L are fixed, unknown constants $\rightarrow$ ~~probability~~
- MLE: $\hat{\theta} = \frac{\text{\# postive reviews}}{\text{total \# reviews}}$

Bayesian vs. Frequentist inference in practice

Frequentist inference

- θ_B and θ_L are fixed, unknown constants $\rightarrow$ ~~probability~~
- MLE: $\hat{\theta} = \frac{\text{\# postive reviews}}{\text{total \# reviews}}$
- Confidence Interval: $\hat{\theta} \pm z_{\alpha/2}\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}$

Bayesian vs. Frequentist inference in practice

Frequentist inference

- θ_B and θ_L are fixed, unknown constants $\rightarrow$ ~~probability~~
- MLE: $\hat{\theta} = \frac{\text{\# postive reviews}}{\text{total \# reviews}}$
- Confidence Interval: $\hat{\theta} \pm z_{\alpha/2}\sqrt{\frac{\hat{\theta}(1-\hat{\theta})}{n}}$

Frequentist inference

Seller	MLE	95% Conf. Interval
BigNBooks	0.967	(0.90, 1.03)
LittleNBooks	1.000	(1.00, 1.00)

Bayesian vs. Frequentist inference in practice

Bayesian inference

- θ_B and θ_L are fixed, unknown constants.

Bayesian vs. Frequentist inference in practice

Bayesian inference

- θ_B and θ_L are fixed, unknown constants.
- A Bayesian uses probability to represent uncertainty about θ_B and θ_L.

Bayesian vs. Frequentist inference in practice

Bayesian inference

- θ_B and θ_L are fixed, unknown constants.
- A Bayesian uses probability to represent uncertainty about θ_B and θ_L.
- A Bayesian must choose prior distributions (often shortened to "priors") for θ_B and θ_L

Bayesian vs. Frequentist inference in practice

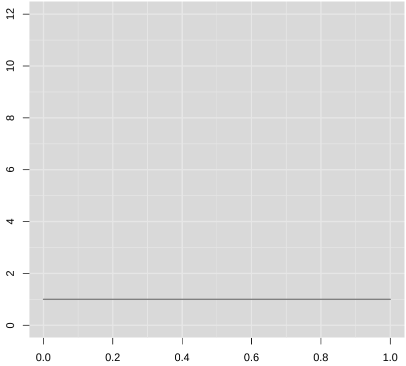
Bayesian inference

- θ_B and θ_L are fixed, unknown constants.
- A Bayesian uses probability to represent uncertainty about θ_B and θ_L.
- A Bayesian must choose prior distributions (often shortened to "priors") for θ_B and θ_L
- For this problem $p(\theta_B) = p(\theta_L) = \text{Unif}(0, 1)$.

Bayesian vs. Frequentist inference in practice

Bayesian inference

- θ_B and θ_L are fixed, unknown constants.
- A Bayesian uses probability to represent uncertainty about θ_B and θ_L.
- A Bayesian must choose prior distributions (often shortened to "priors") for θ_B and θ_L
- For this problem $p(\theta_B) = p(\theta_L) = \text{Unif}(0, 1)$.



Bayesian vs. Frequentist inference in practice

The posterior distribution

Consider θ_B (things are similar for θ_L)

- $p(\theta_B)$ represents our uncertainty about θ_B before observing the data. The Bayesian wants

$$p(\theta_B|\mathbf{Y}^{(B)})$$

the posterior distribution.

Bayesian vs. Frequentist inference in practice

The posterior distribution

Consider θ_B (things are similar for θ_L)

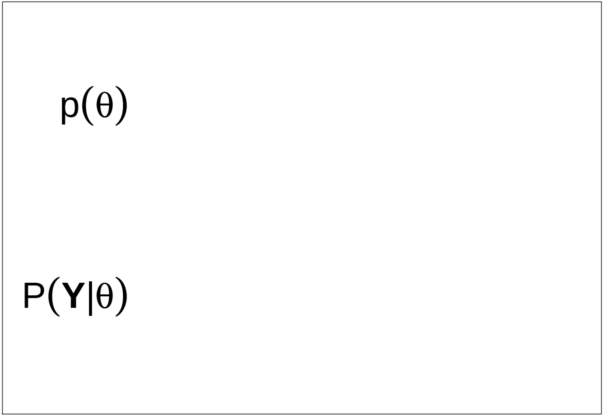
- $p(\theta_B)$ represents our uncertainty about θ_B before observing the data. The Bayesian wants
$$p(\theta_B|\mathbf{Y}^{(B)})$$
the posterior distribution.
- The posterior distribution represents our uncertainty about θ_B after observing the data.

Bayesian vs. Frequentist inference in practice

Bayes Theorem

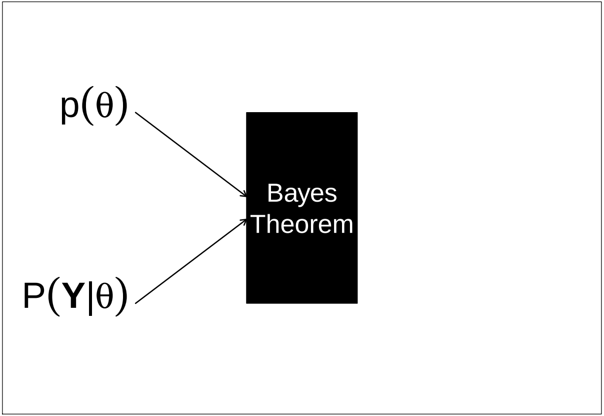
Bayesian vs. Frequentist inference in practice

Bayes Theorem



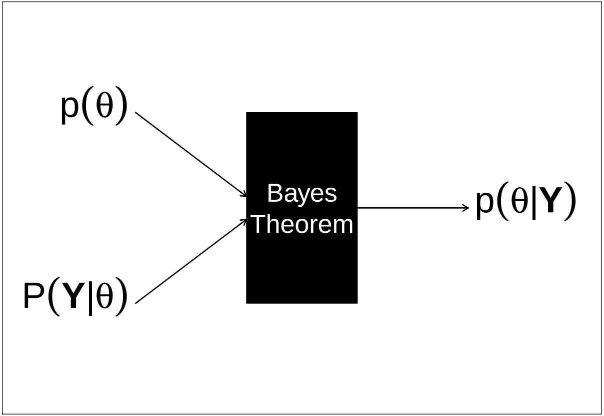
Bayesian vs. Frequentist inference in practice

Bayes Theorem



Bayesian vs. Frequentist inference in practice

Bayes Theorem



Bayesian vs. Frequentist inference in practice

Posterior distributions for θ_B and θ_L

Posterior distributions

Seller	Prior	Likelihood	Posterior
BigNBooks			

Bayesian vs. Frequentist inference in practice

Posterior distributions for θ_B and θ_L

Posterior distributions

Seller	Prior	Likelihood	Posterior
BigNBooks	$\text{Unif}(0, 1)$		

Bayesian vs. Frequentist inference in practice

Posterior distributions for θ_B and θ_L

Posterior distributions

Seller	Prior	Likelihood	Posterior
BigNBooks	Unif$(0,1)$	$\theta_B^{29}(1-\theta_L)^1$	

Bayesian vs. Frequentist inference in practice

Posterior distributions for θ_B and θ_L

Posterior distributions

Seller	Prior	Likelihood	Posterior
BigNBooks	Unif$(0,1)$	$\theta_B^{29}(1-\theta_L)^1$	Beta$(29+1,1+1)$

Bayesian vs. Frequentist inference in practice

Posterior distributions for θ_B and θ_L

Posterior distributions

Seller	Prior	Likelihood	Posterior
BigNBooks	Unif$(0, 1)$	$\theta_B^{29}(1 - \theta_L)^1$	Beta$(29 + 1, 1 + 1)$
LittleNBooks			

Bayesian vs. Frequentist inference in practice

Posterior distributions for θ_B and θ_L

Posterior distributions

Seller	Prior	Likelihood	Posterior
BigNBooks	Unif$(0,1)$	$\theta_B^{29}(1-\theta_L)^1$	Beta$(29+1, 1+1)$
LittleNBooks	Unif$(0,1)$		

Bayesian vs. Frequentist inference in practice

Posterior distributions for θ_B and θ_L

Posterior distributions

Seller	Prior	Likelihood	Posterior
BigNBooks	Unif$(0,1)$	$\theta_B^{29}(1-\theta_L)^1$	Beta$(29+1,1+1)$
LittleNBooks	Unif$(0,1)$	$\theta_L^{5}(1-\theta_L)^0$	

Bayesian vs. Frequentist inference in practice

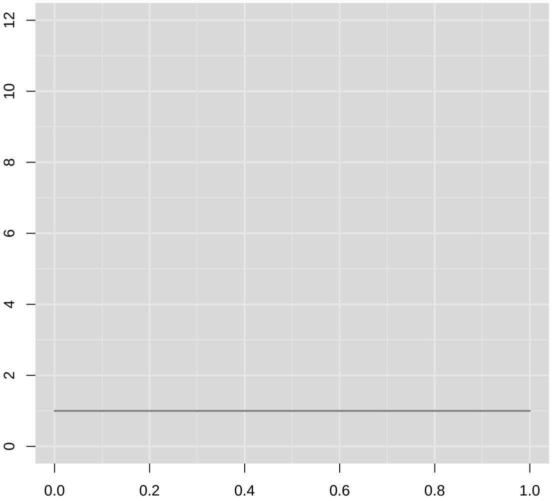
Posterior distributions for θ_B and θ_L

Posterior distributions

Seller	Prior	Likelihood	Posterior
BigNBooks	Unif$(0, 1)$	$\theta_B^{29}(1 - \theta_L)^1$	Beta$(29 + 1, 1 + 1)$
LittleNBooks	Unif$(0, 1)$	$\theta_L^5(1 - \theta_L)^0$	Beta$(5 + 1, 0 + 1)$

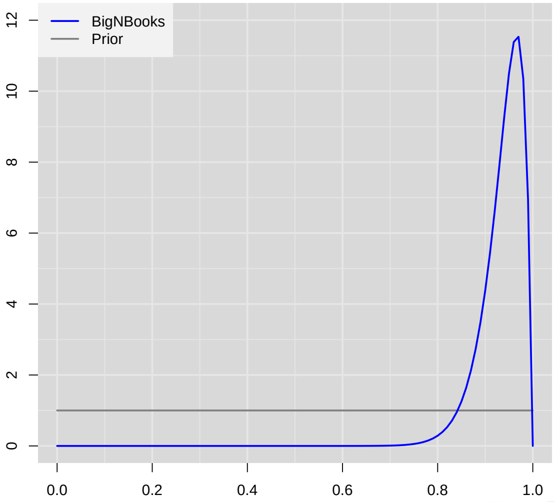
Bayesian vs. Frequentist inference in practice

Posterior distributions for θ_B and θ_L



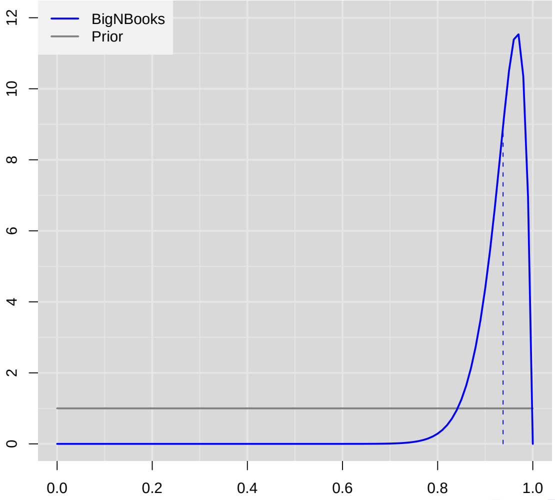
Bayesian vs. Frequentist inference in practice

Posterior distributions for θ_B and θ_L



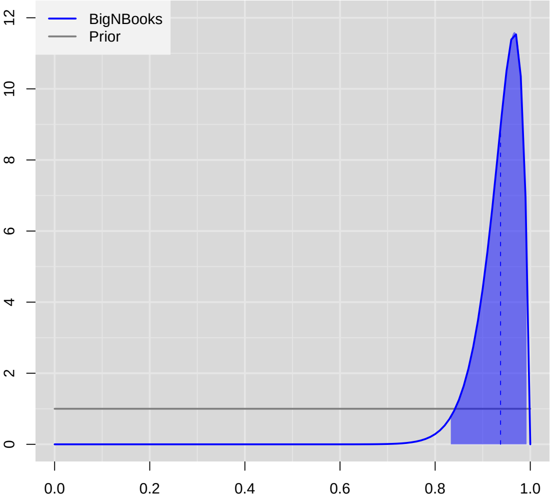
Bayesian vs. Frequentist inference in practice

Posterior distributions for θ_B and θ_L



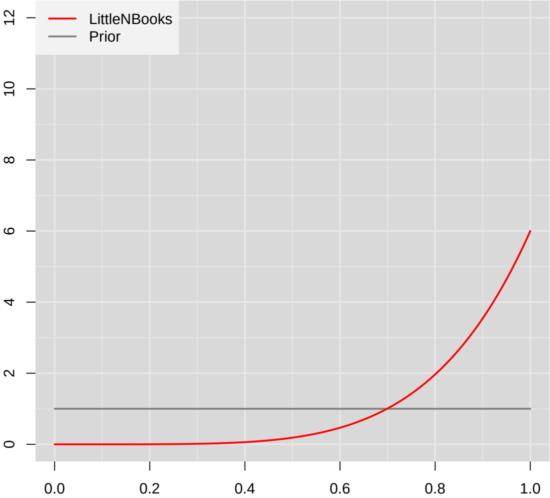
Bayesian vs. Frequentist inference in practice

Posterior distributions for θ_B and θ_L



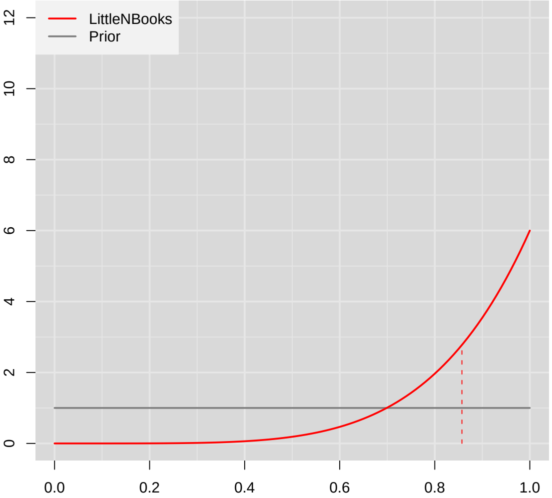
Bayesian vs. Frequentist inference in practice

Posterior distributions for θ_B and θ_L



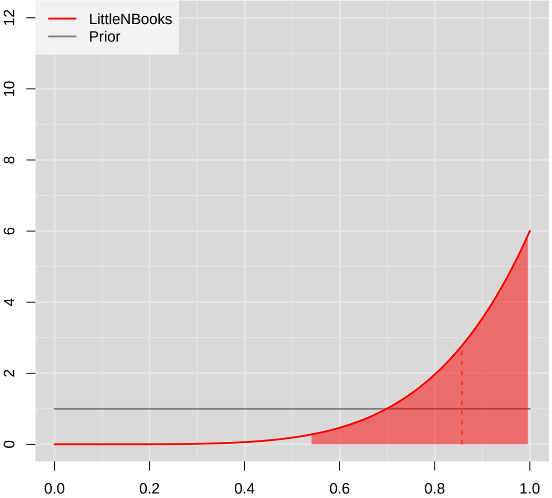
Bayesian vs. Frequentist inference in practice

Posterior distributions for θ_B and θ_L



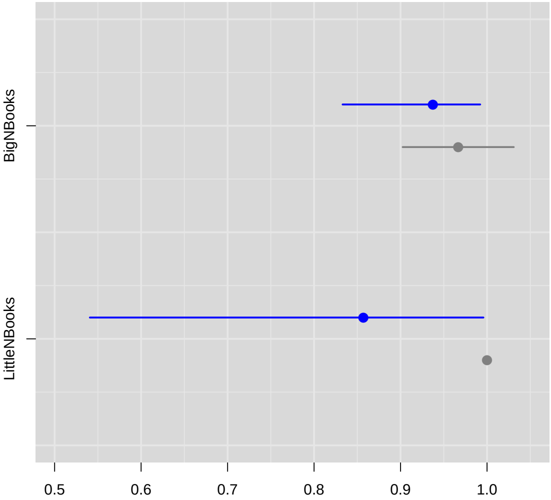
Bayesian vs. Frequentist inference in practice

Posterior distributions for θ_B and θ_L



Bayesian vs. Frequentist inference in practice

Comparing the estimates



Outline

Thinking about tests

- Test takers
- Test items

Thinking about tests

- Test takers
  - Test takers have different levels of ability.
- Test items

Thinking about tests

- Test takers
  - Test takers have different levels of ability.
- Test items
  - Some test items are more difficult than others.

Thinking about tests

- Test takers
  - Test takers have different levels of ability.
- Test items
  - Some test items are more difficult than others.
  - Some test items are better (more "discriminating") than others.

A mathematical model for tests

Defining Greek symbols

- Consider a test with p items ($j = 1, \ldots, p$).

A mathematical model for tests

Defining Greek symbols

- Consider a test with p items ($j = 1, \ldots, p$).
- Let δ_j be the difficulty of item j.

A mathematical model for tests

Defining Greek symbols

- Consider a test with p items ($j = 1, \ldots, p$).
- Let δ_j be the difficulty of item j.
- Let α_j be the discrimination of item j.

A mathematical model for tests

Defining Greek symbols

- Consider a test with p items ($j = 1, \ldots, p$).
- Let δ_j be the difficulty of item j.
- Let α_j be the discrimination of item j.
- Let

$$Y_j = \begin{cases} 1 & \text{if an individual endorses } j\text{-th item} \\ 0 & \text{otherwise} \end{cases}$$

A mathematical model for tests

Defining Greek symbols

- Consider a test with p items ($j = 1, \ldots, p$).
- Let δ_j be the difficulty of item j.
- Let α_j be the discrimination of item j.
- Let

$$Y_j = \begin{cases} 1 & \text{if an individual endorses } j\text{-th item} \\ 0 & \text{otherwise} \end{cases}$$

- Let θ be the ability of an individual test taker.

A mathematical model for tests

The probability of success on j^{th} item

A mathematical model for tests

The probability of success on j^{th} item

$$P(Y_j = 1|\theta) = \frac{1}{1 + e^{-\alpha_j(\theta - \delta_j)}}$$

A mathematical model for tests

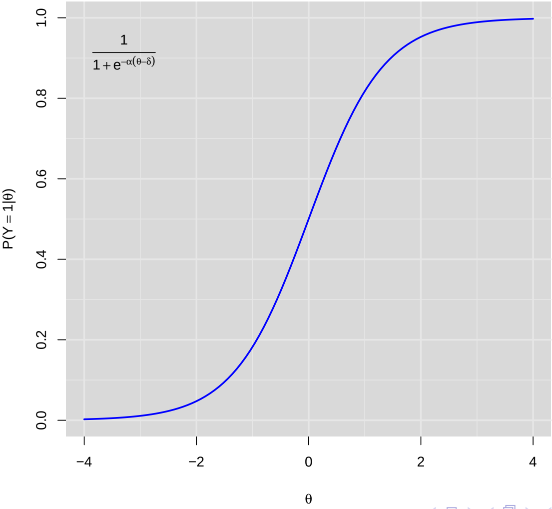
The probability of success on j^{th} item

$$P(Y_j = 1|\theta) = \frac{1}{1 + e^{-\alpha_j(\theta - \delta_j)}}$$

$$P(Y_j = 0|\theta) = 1 - \frac{1}{1 + e^{-\alpha_j(\theta - \delta_j)}}$$

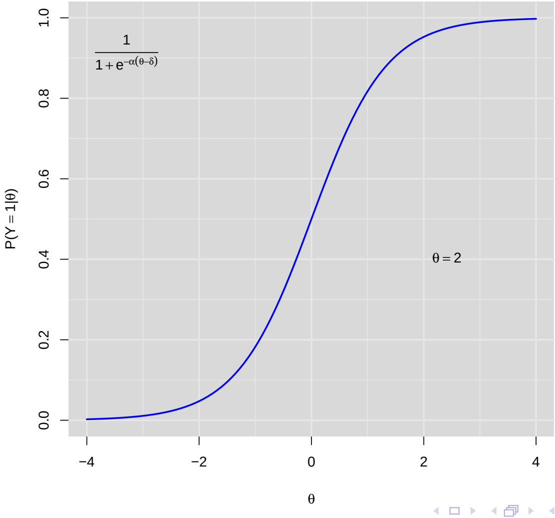
Visualizing the model

An item with "average" difficulty $\delta_j = 0.0$ ($\alpha_j = 1.5$)



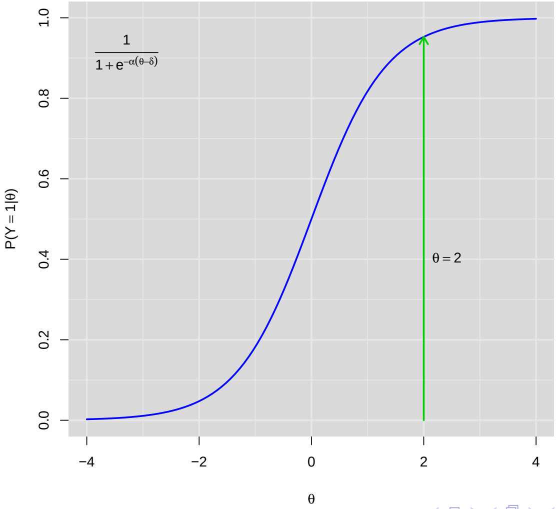
Visualizing the model

An item with "average" difficulty $\delta_j = 0.0$ ($\alpha_j = 1.5$)



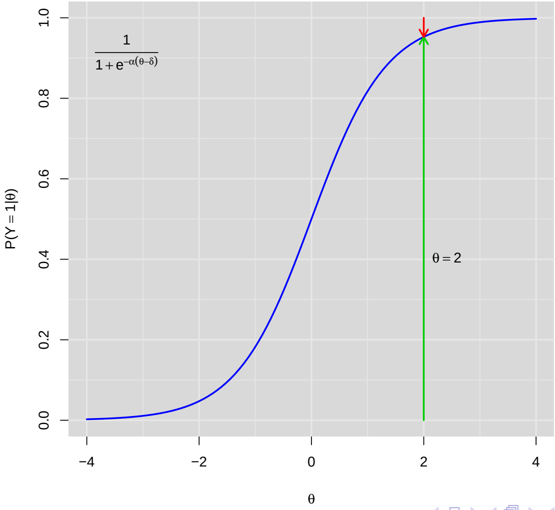
Visualizing the model

An item with “average” difficulty $\delta_j = 0.0$ ($\alpha_j = 1.5$)



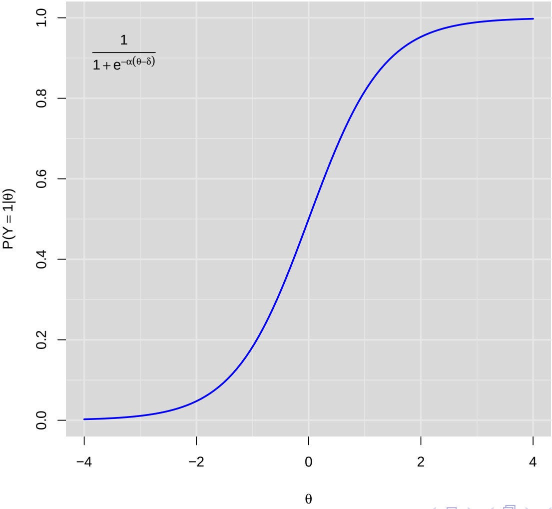
Visualizing the model

An item with “average” difficulty $\delta_j = 0.0$ ($\alpha_j = 1.5$)



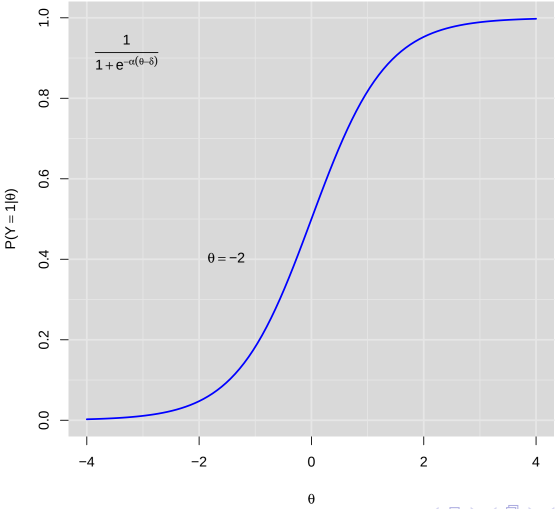
Visualizing the model

An item with "average" difficulty $\delta_j = 0.0$ ($\alpha_j = 1.5$)



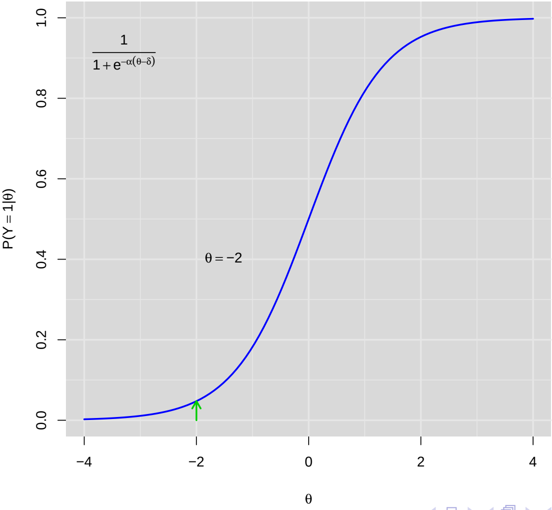
Visualizing the model

An item with "average" difficulty $\delta_j = 0.0$ ($\alpha_j = 1.5$)



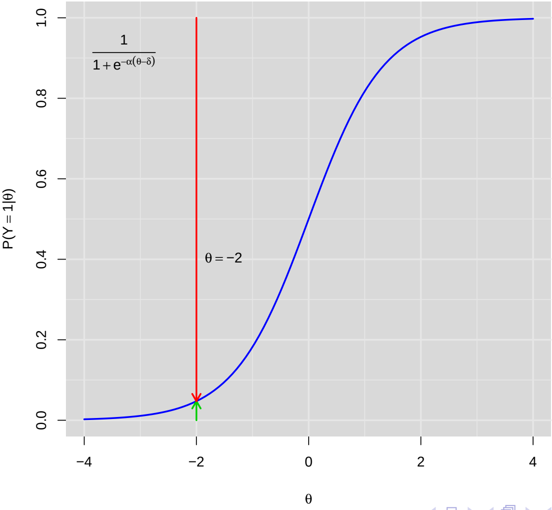
Visualizing the model

An item with "average" difficulty $\delta_j = 0.0$ ($\alpha_j = 1.5$)



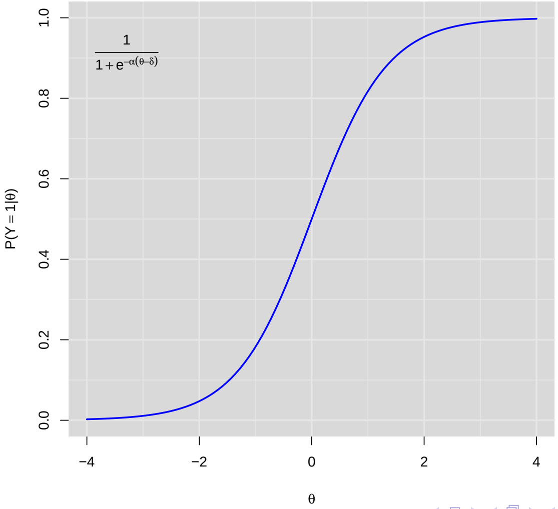
Visualizing the model

An item with "average" difficulty $\delta_j = 0.0$ ($\alpha_j = 1.5$)



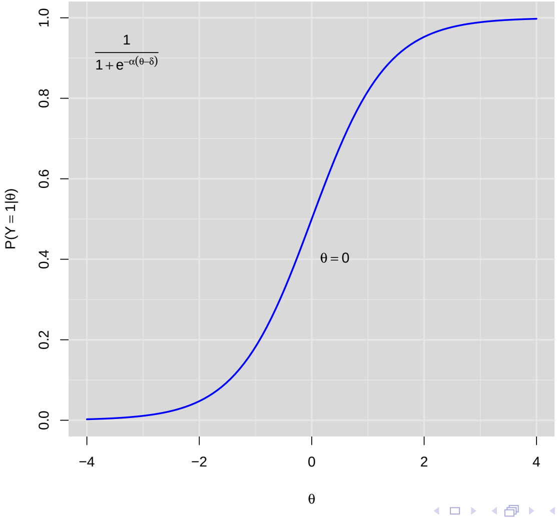
Visualizing the model

An item with "average" difficulty $\delta_j = 0.0$ ($\alpha_j = 1.5$)



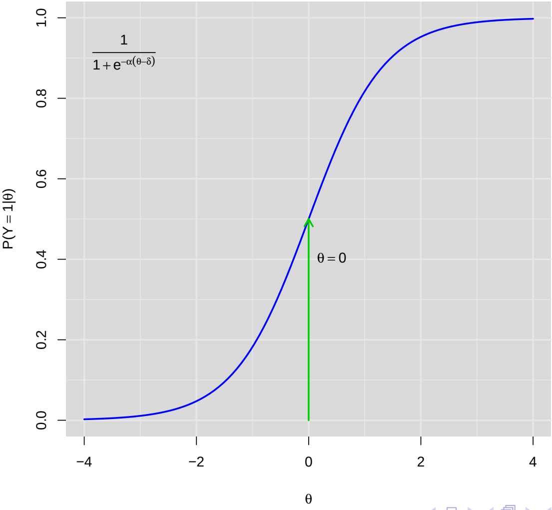
Visualizing the model

An item with "average" difficulty $\delta_j = 0.0$ ($\alpha_j = 1.5$)



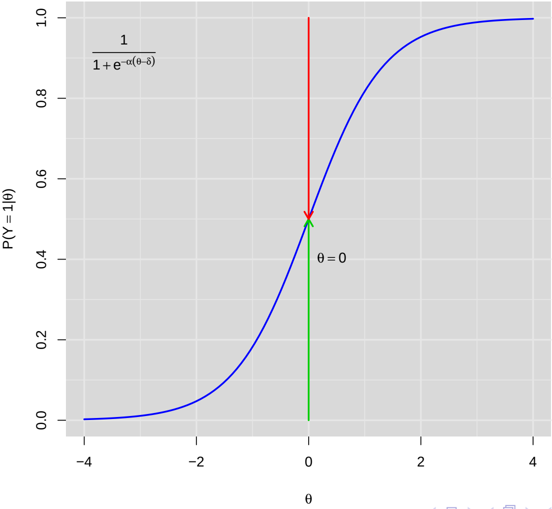
Visualizing the model

An item with "average" difficulty $\delta_j = 0.0$ ($\alpha_j = 1.5$)



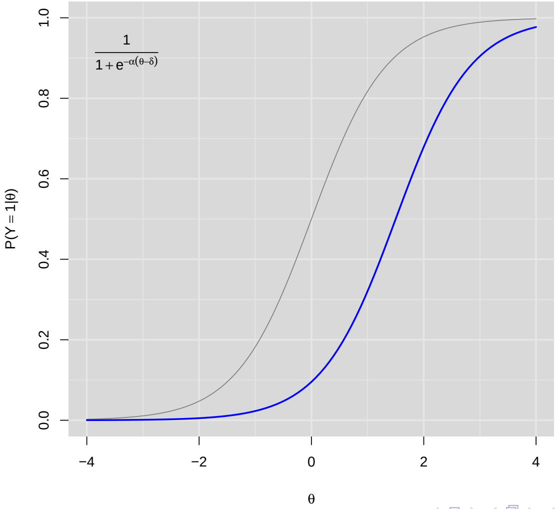
Visualizing the model

An item with “average” difficulty $\delta_j = 0.0$ ($\alpha_j = 1.5$)



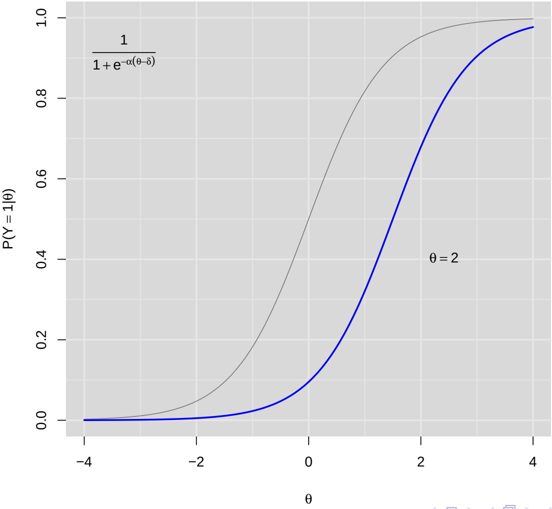
Visualizing the model

An item with above-average difficulty $\delta_j = 1.5$ ($\alpha_j = 1.5$)



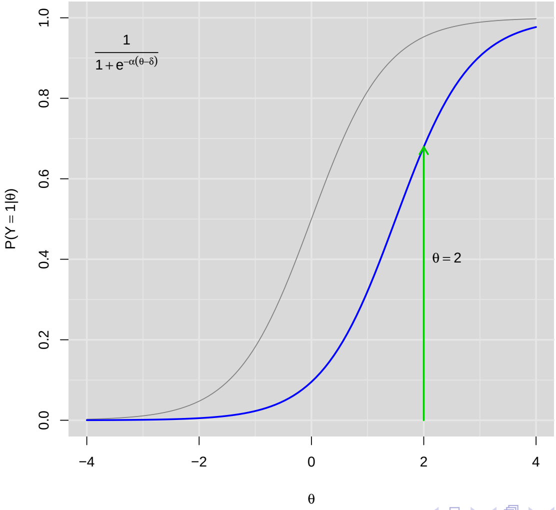
Visualizing the model

An item with above-average difficulty $\delta_j = 1.5$ ($\alpha_j = 1.5$)



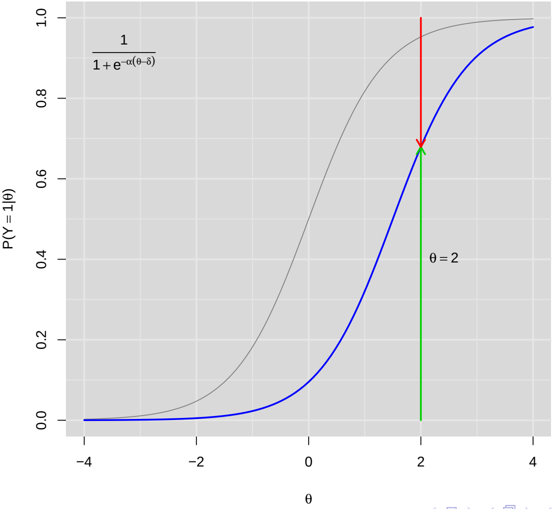
Visualizing the model

An item with above-average difficulty $\delta_j = 1.5$ $(\alpha_j = 1.5)$



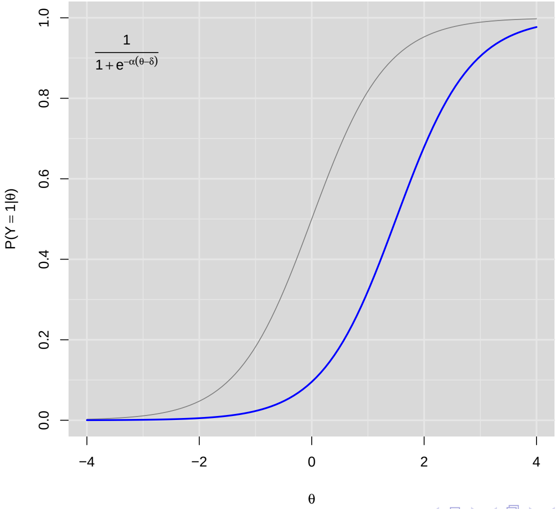
Visualizing the model

An item with above-average difficulty $\delta_j = 1.5$ ($\alpha_j = 1.5$)



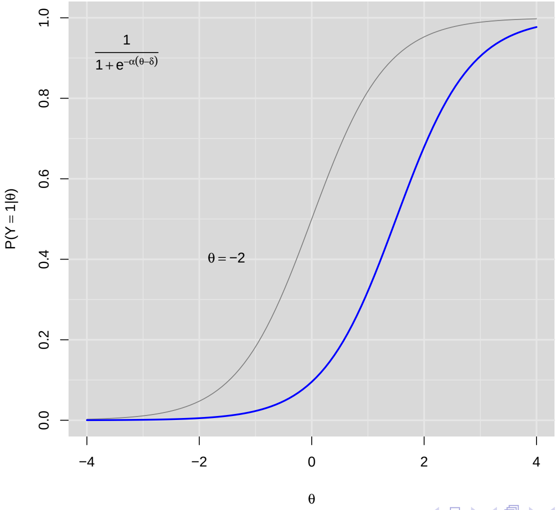
Visualizing the model

An item with above-average difficulty $\delta_j = 1.5$ ($\alpha_j = 1.5$)



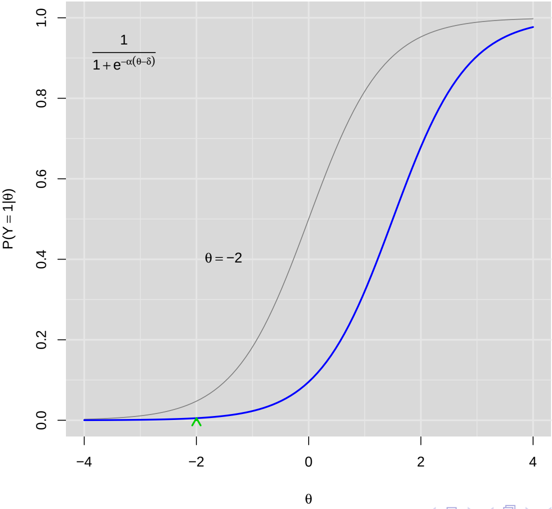
Visualizing the model

An item with above-average difficulty $\delta_j = 1.5$ ($\alpha_j = 1.5$)



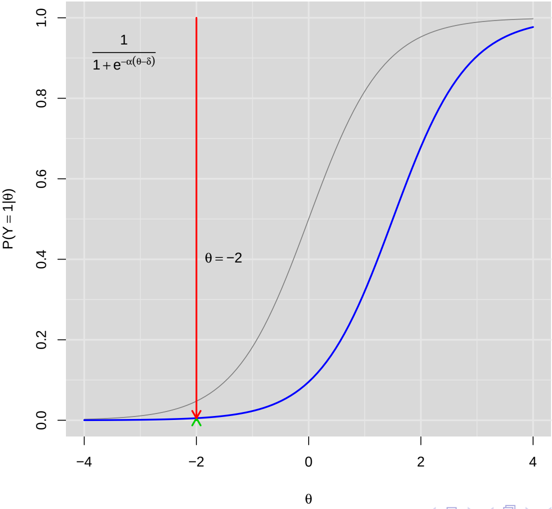
Visualizing the model

An item with above-average difficulty $\delta_j = 1.5$ ($\alpha_j = 1.5$)



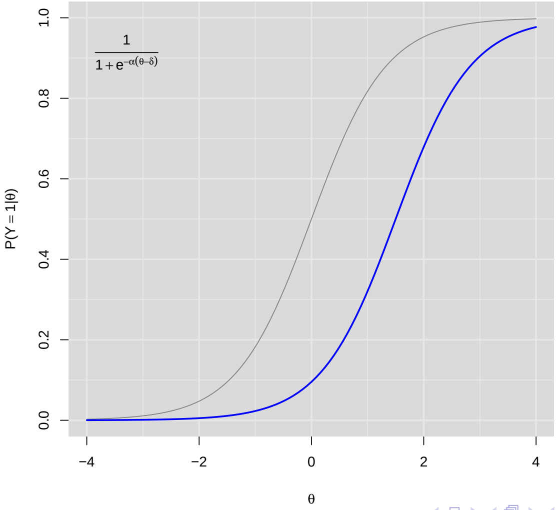
Visualizing the model

An item with above-average difficulty $\delta_j = 1.5$ ($\alpha_j = 1.5$)



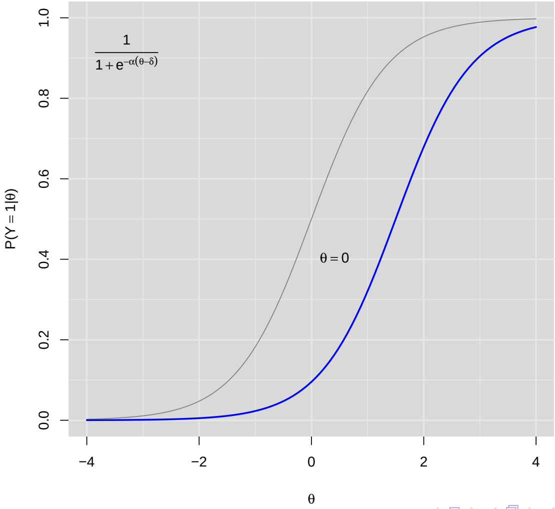
Visualizing the model

An item with above-average difficulty $\delta_j = 1.5$ ($\alpha_j = 1.5$)



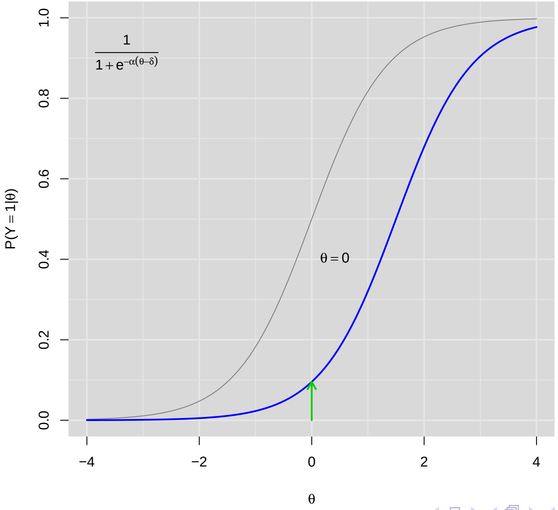
Visualizing the model

An item with above-average difficulty $\delta_j = 1.5$ ($\alpha_j = 1.5$)



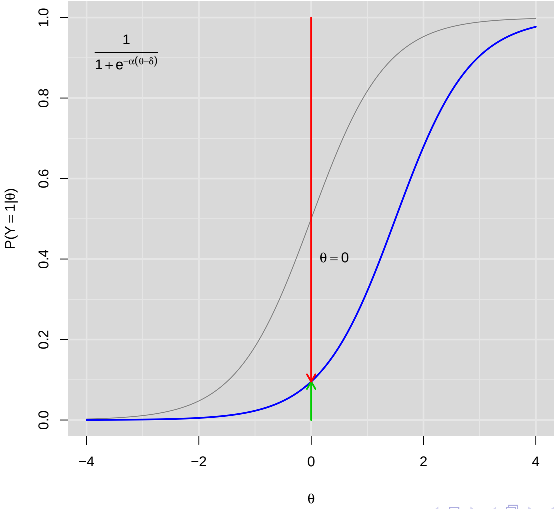
Visualizing the model

An item with above-average difficulty $\delta_j = 1.5$ ($\alpha_j = 1.5$)



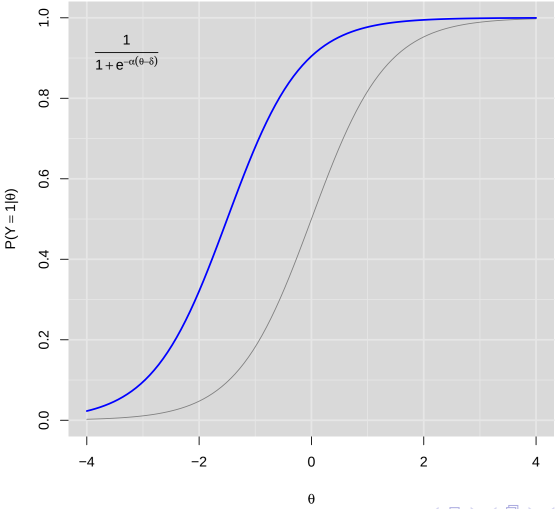
Visualizing the model

An item with above-average difficulty $\delta_j = 1.5$ ($\alpha_j = 1.5$)



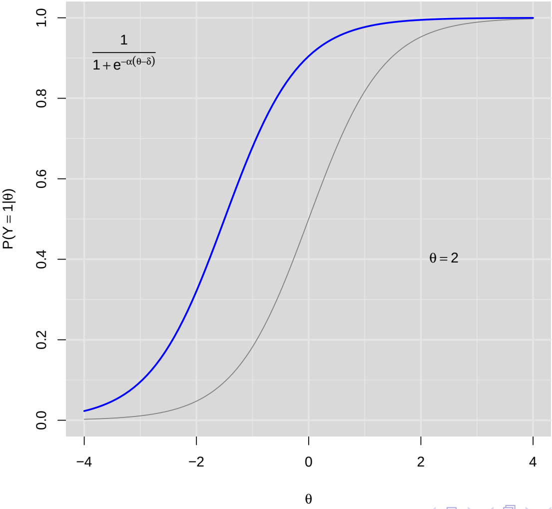
Visualizing the model

An item with below-average difficulty $\delta_j = -1.5$ ($\alpha_j = 1.5$)



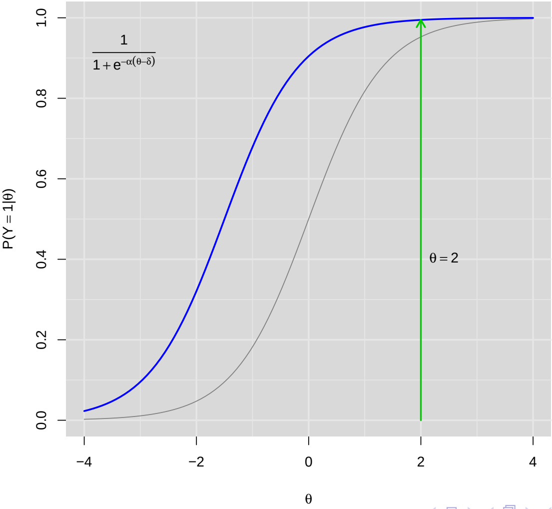
Visualizing the model

An item with below-average difficulty $\delta_j = -1.5$ ($\alpha_j = 1.5$)



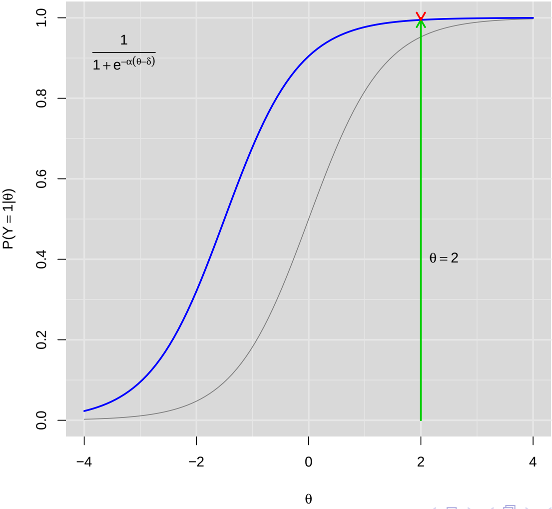
Visualizing the model

An item with below-average difficulty $\delta_j = -1.5$ ($\alpha_j = 1.5$)



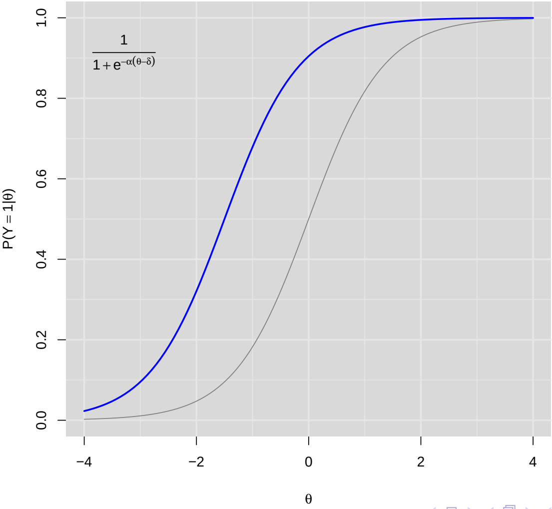
Visualizing the model

An item with below-average difficulty $\delta_j = -1.5$ ($\alpha_j = 1.5$)



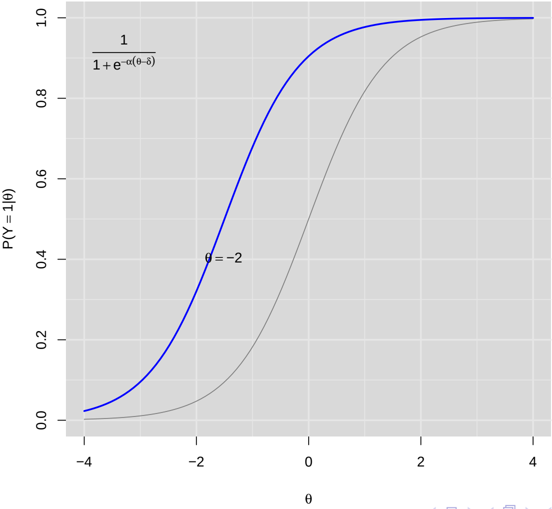
Visualizing the model

An item with below-average difficulty $\delta_j = -1.5$ ($\alpha_j = 1.5$)



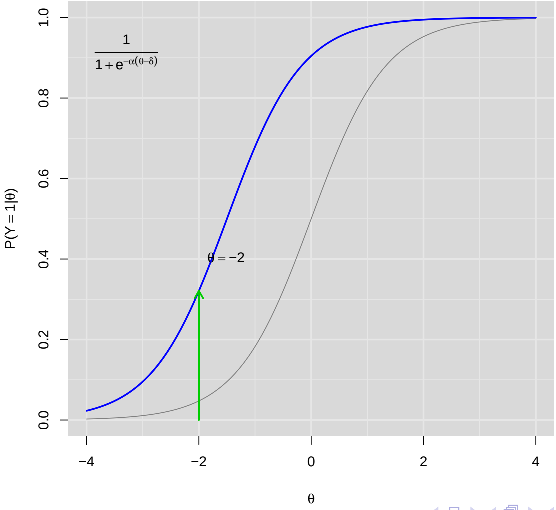
Visualizing the model

An item with below-average difficulty $\delta_j = -1.5$ ($\alpha_j = 1.5$)



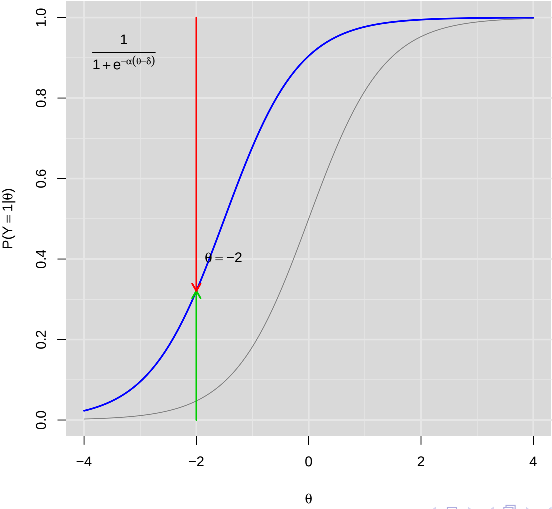
Visualizing the model

An item with below-average difficulty $\delta_j = -1.5$ ($\alpha_j = 1.5$)



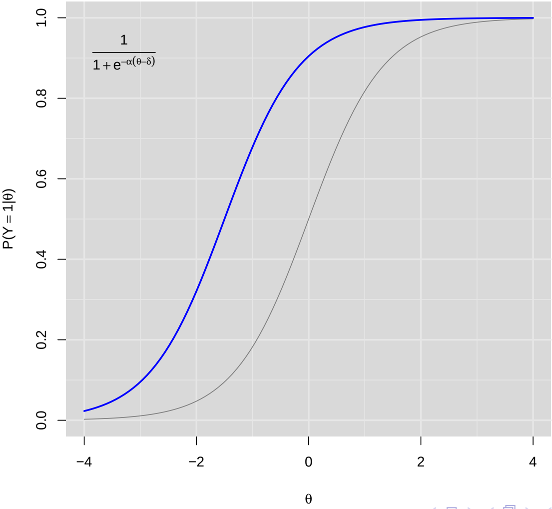
Visualizing the model

An item with below-average difficulty $\delta_j = -1.5$ ($\alpha_j = 1.5$)



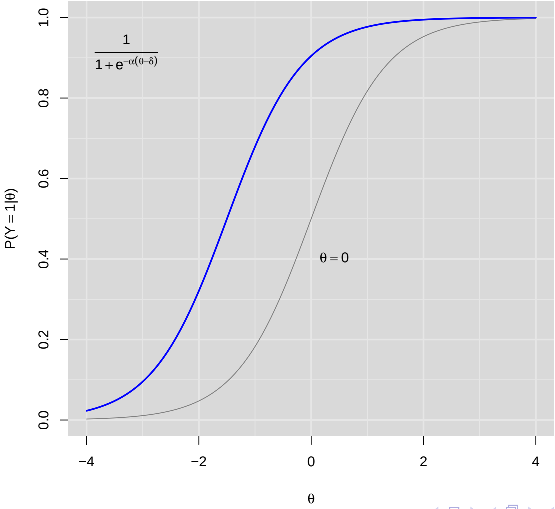
Visualizing the model

An item with below-average difficulty $\delta_j = -1.5$ ($\alpha_j = 1.5$)



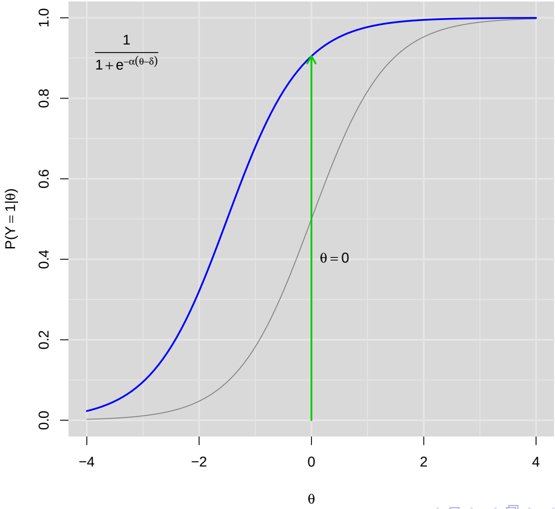
Visualizing the model

An item with below-average difficulty $\delta_j = -1.5$ ($\alpha_j = 1.5$)



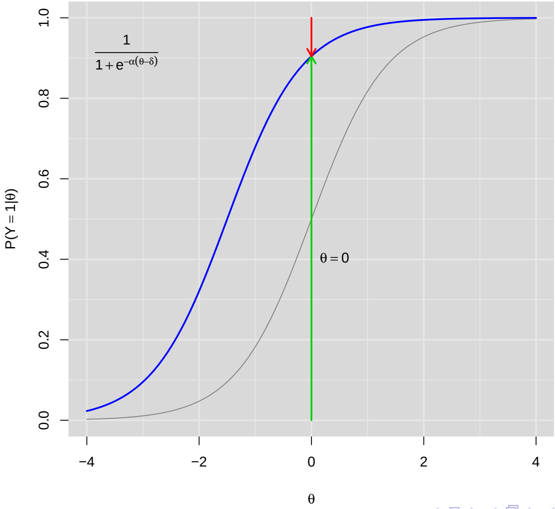
Visualizing the model

An item with below-average difficulty $\delta_j = -1.5$ ($\alpha_j = 1.5$)



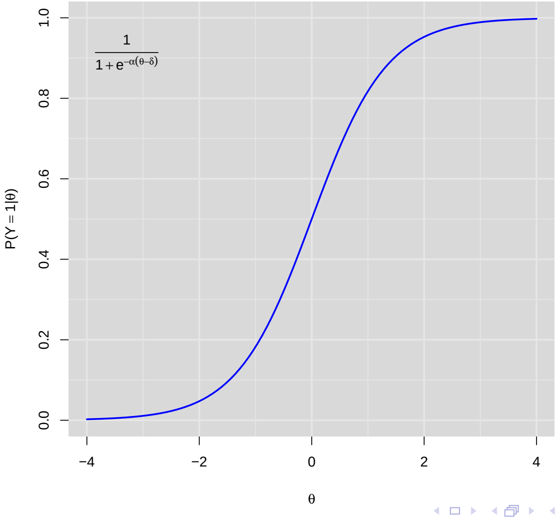
Visualizing the model

An item with below-average difficulty $\delta_j = -1.5$ ($\alpha_j = 1.5$)



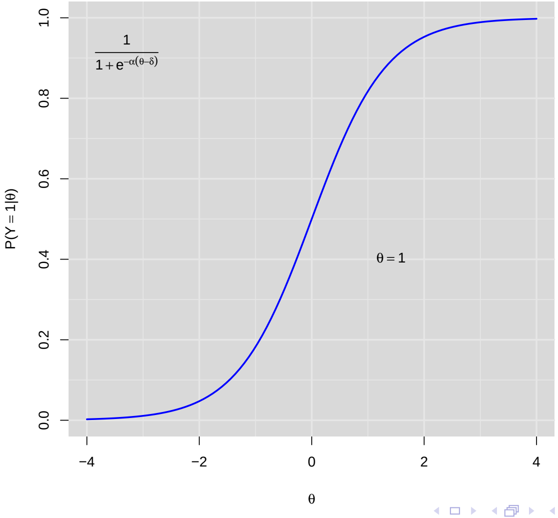
Visualizing the model

An item with average discrimination $\alpha_j = 1.5$ ($\delta_j = 0.0$)



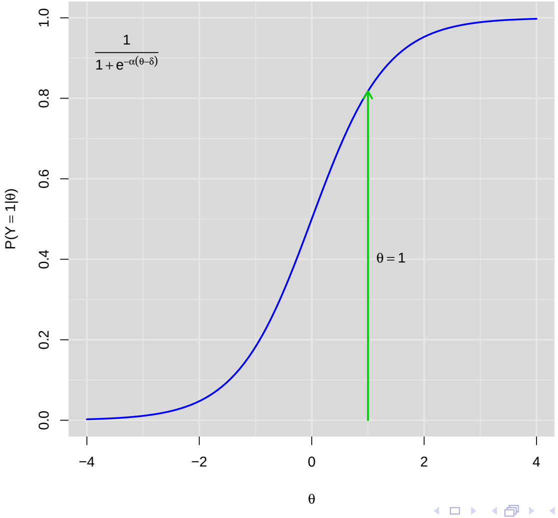
Visualizing the model

An item with average discrimination $\alpha_j = 1.5$ ($\delta_j = 0.0$)



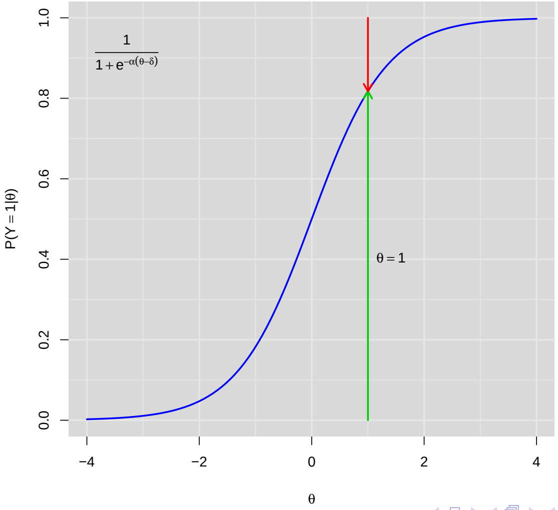
Visualizing the model

An item with average discrimination $\alpha_j = 1.5$ ($\delta_j = 0.0$)



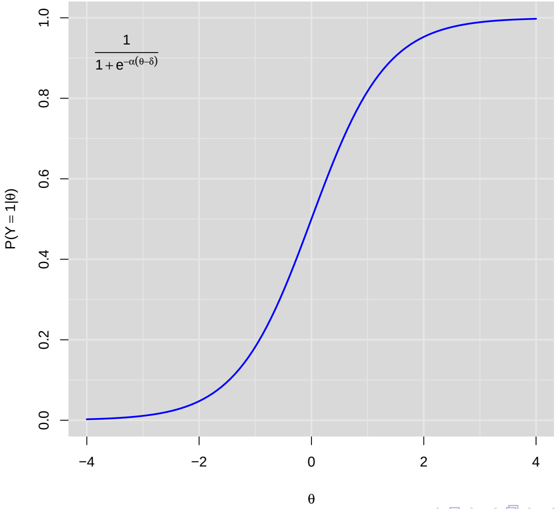
Visualizing the model

An item with average discrimination $\alpha_j = 1.5$ ($\delta_j = 0.0$)



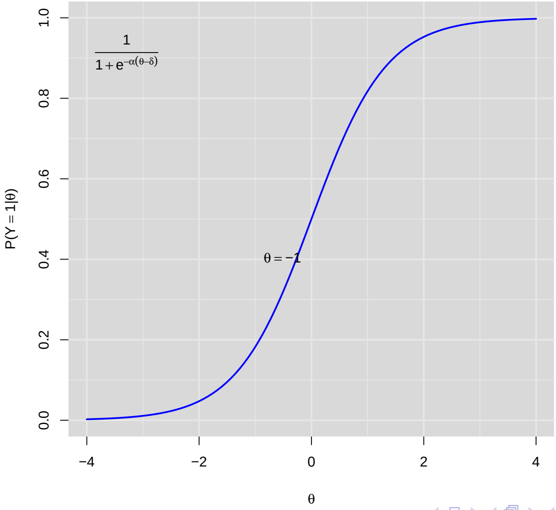
Visualizing the model

An item with average discrimination $\alpha_j = 1.5$ ($\delta_j = 0.0$)



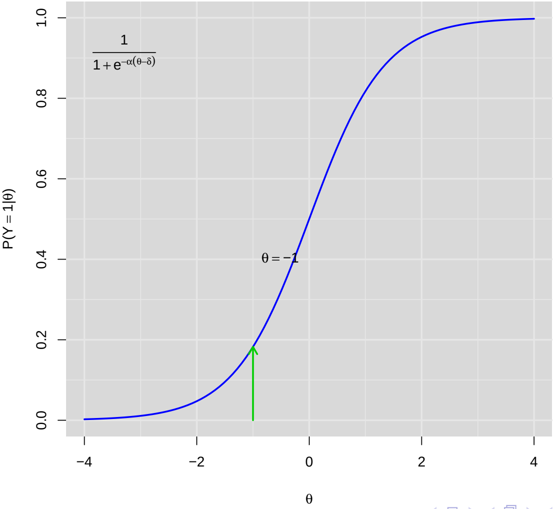
Visualizing the model

An item with average discrimination $\alpha_j = 1.5$ ($\delta_j = 0.0$)



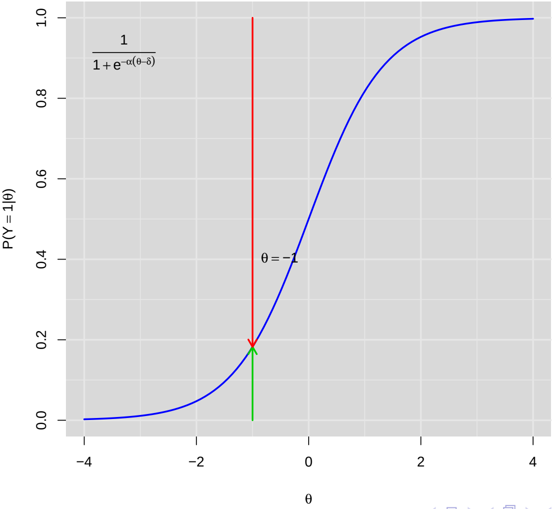
Visualizing the model

An item with average discrimination $\alpha_j = 1.5$ ($\delta_j = 0.0$)



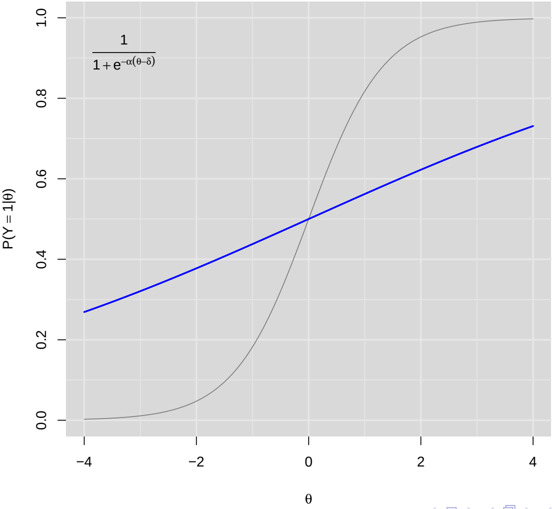
Visualizing the model

An item with average discrimination $\alpha_j = 1.5$ ($\delta_j = 0.0$)



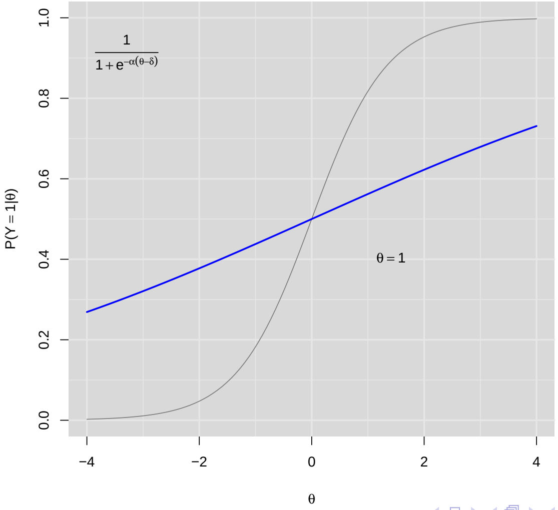
Visualizing the model

An item with below average discrimination $\alpha_j = 0.25$ ($\delta_j = 0.0$)



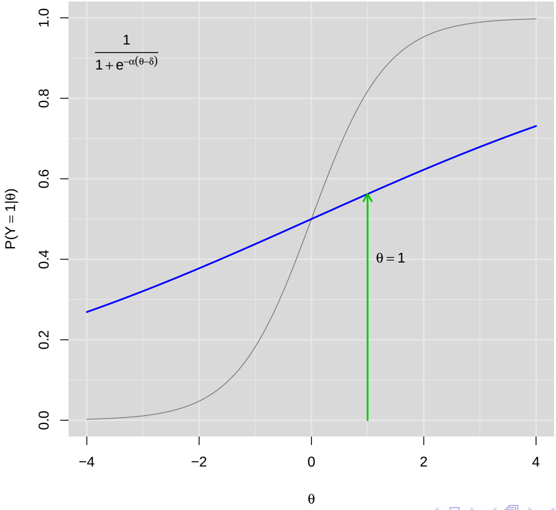
Visualizing the model

An item with below average discrimination $\alpha_j = 0.25$ ($\delta_j = 0.0$)



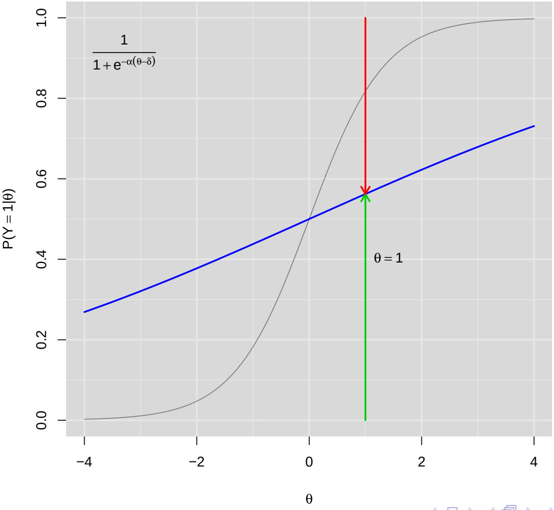
Visualizing the model

An item with below average discrimination $\alpha_j = 0.25$ ($\delta_j = 0.0$)



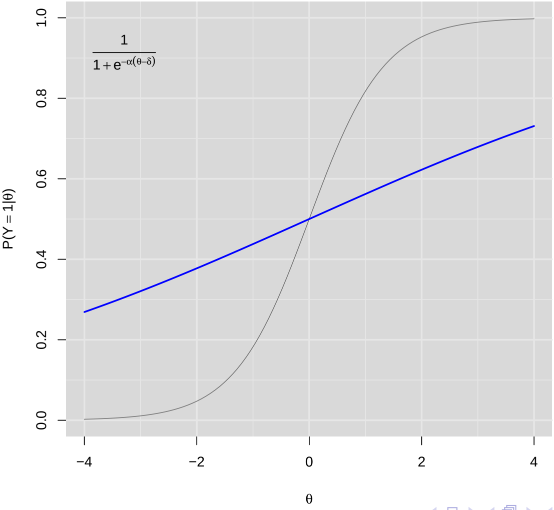
Visualizing the model

An item with below average discrimination $\alpha_j = 0.25$ ($\delta_j = 0.0$)



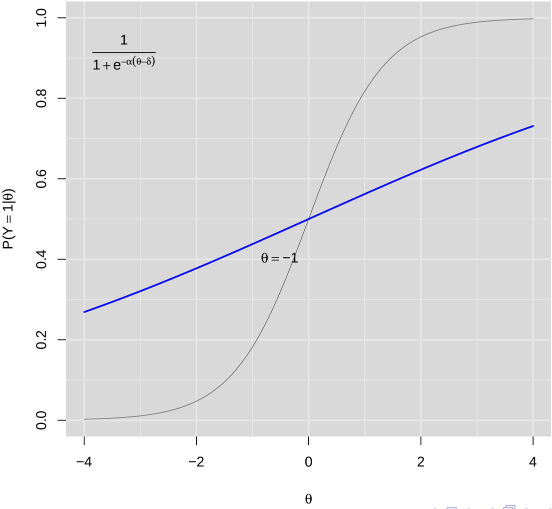
Visualizing the model

An item with below average discrimination $\alpha_j = 0.25$ ($\delta_j = 0.0$)



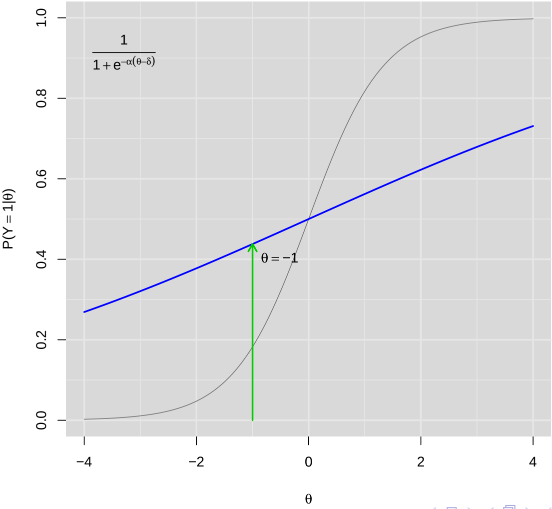
Visualizing the model

An item with below average discrimination $\alpha_j = 0.25$ ($\delta_j = 0.0$)



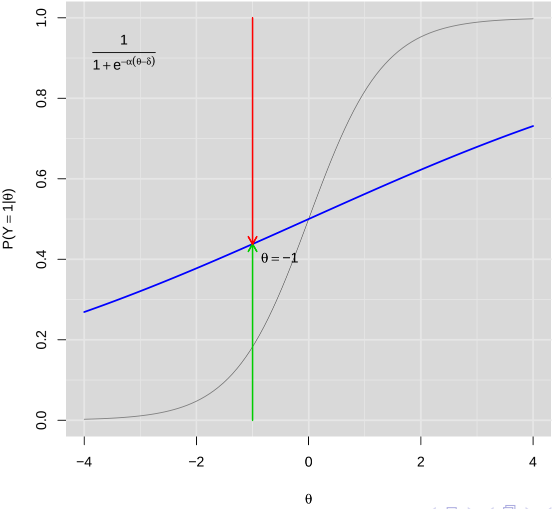
Visualizing the model

An item with below average discrimination $\alpha_j = 0.25$ ($\delta_j = 0.0$)



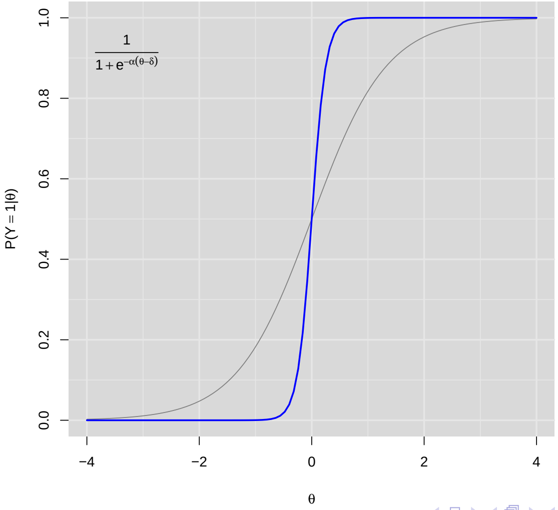
Visualizing the model

An item with below average discrimination $\alpha_j = 0.25$ ($\delta_j = 0.0$)



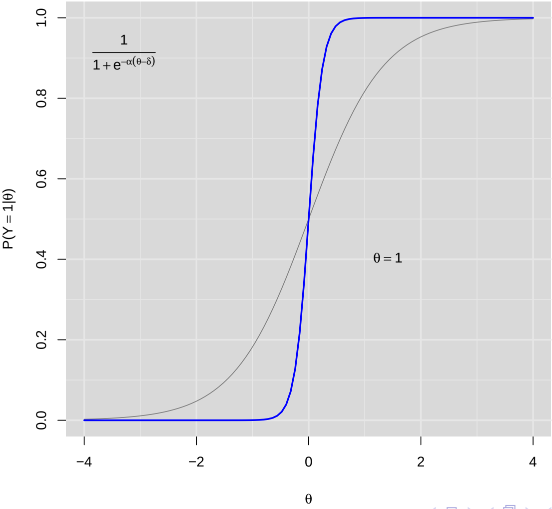
Visualizing the model

An item with above average discrimination $\alpha_j = 8.0$ ($\delta_j = 0.0$)



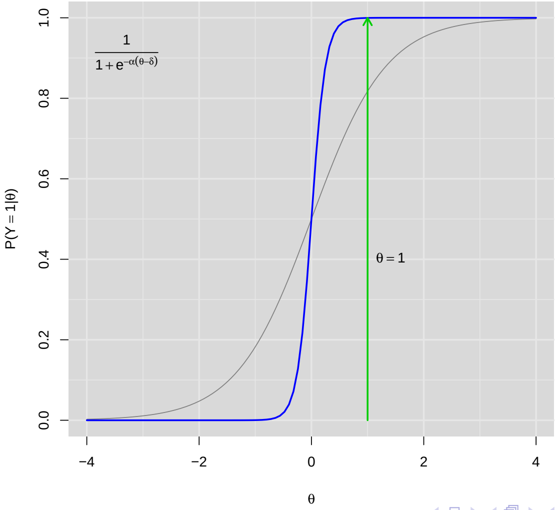
Visualizing the model

An item with above average discrimination $\alpha_j = 8.0$ ($\delta_j = 0.0$)



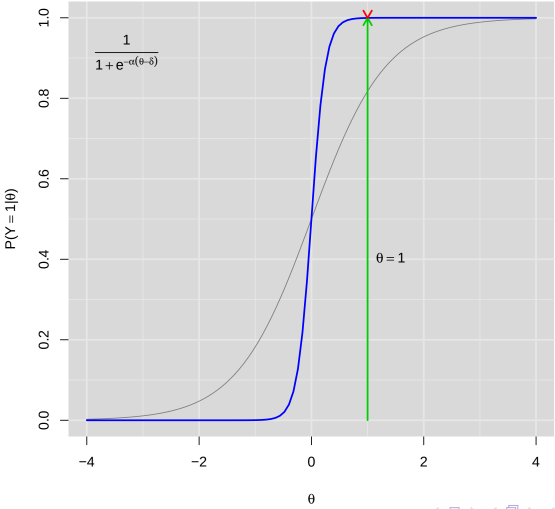
Visualizing the model

An item with above average discrimination $\alpha_j = 8.0$ ($\delta_j = 0.0$)



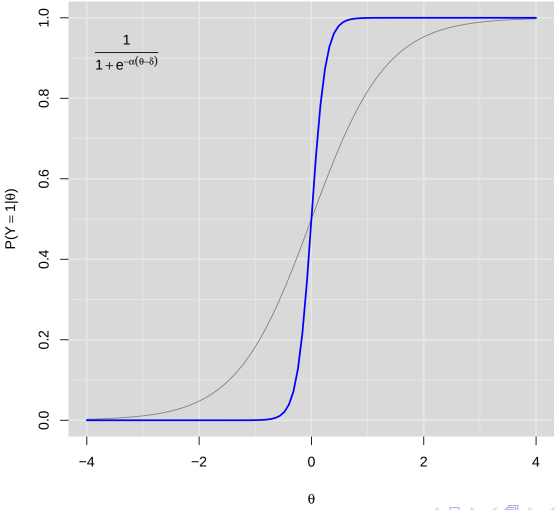
Visualizing the model

An item with above average discrimination $\alpha_j = 8.0$ ($\delta_j = 0.0$)



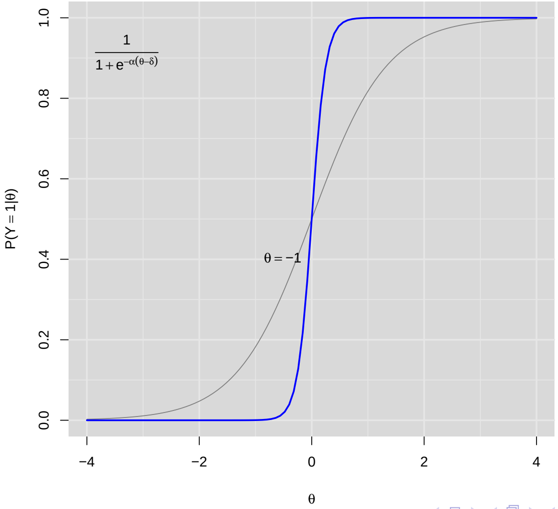
Visualizing the model

An item with above average discrimination $\alpha_j = 8.0$ ($\delta_j = 0.0$)



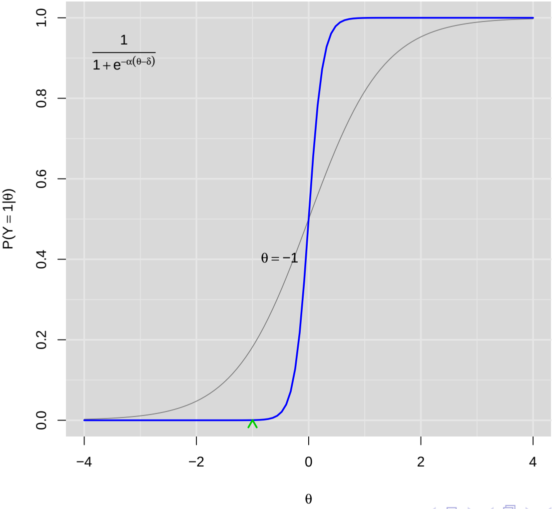
Visualizing the model

An item with above average discrimination $\alpha_j = 8.0$ ($\delta_j = 0.0$)



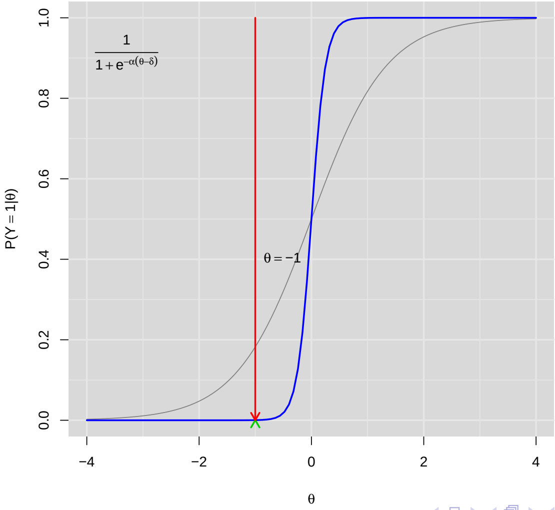
Visualizing the model

An item with above average discrimination $\alpha_j = 8.0$ ($\delta_j = 0.0$)



Visualizing the model

An item with above average discrimination $\alpha_j = 8.0$ ($\delta_j = 0.0$)



Information

A benefit of using an IRT model

- Loosely: Information is measure of how precisely we can estimate some quantity of interest (like θ).

Information

A benefit of using an IRT model

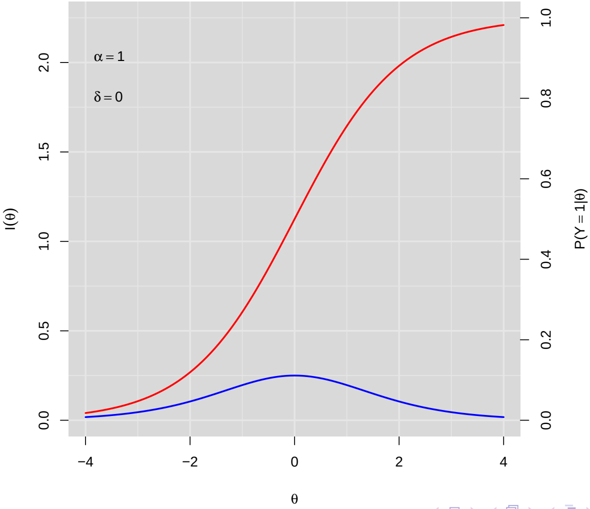
- Loosely: Information is measure of how precisely we can estimate some quantity of interest (like θ).
- Precisely: If $\hat{\theta}$ is the MLE of θ, then

$$I(\theta) = 1/V_\theta(\hat{\theta})$$

where $V_\theta(\hat{\theta})$ is the (asymptotic) variance of the MLE $\hat{\theta}$.

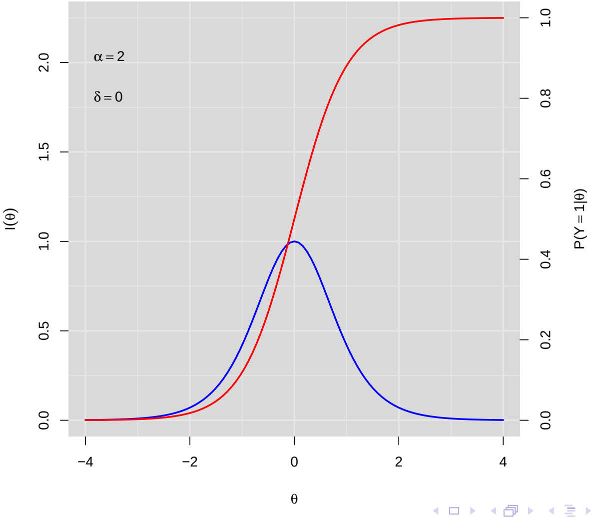
Information

Item information curves



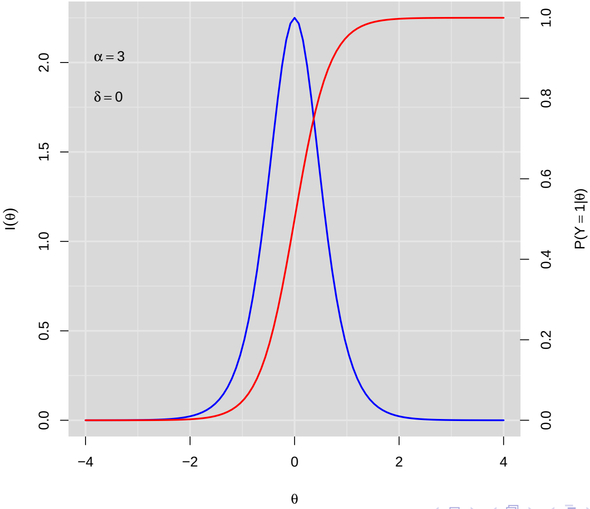
Information

Item information curves



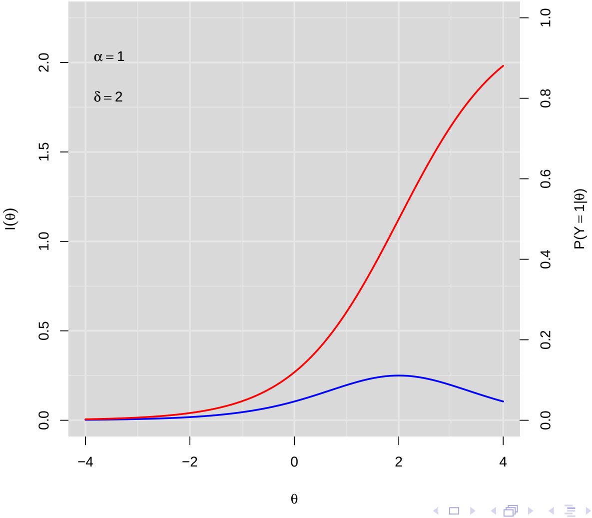
Information

Item information curves



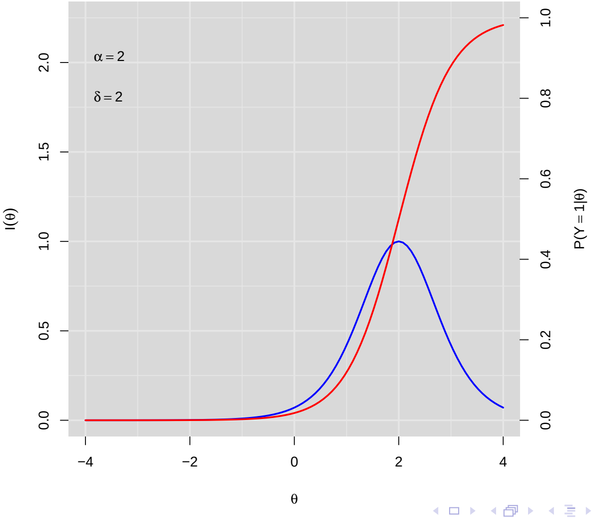
Information

Item information curves



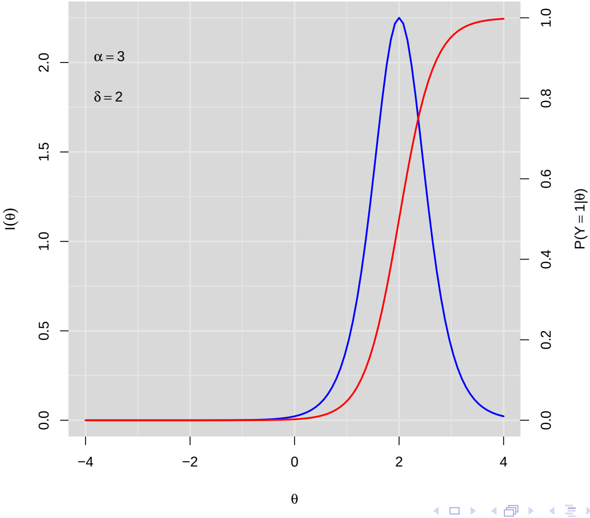
Information

Item information curves



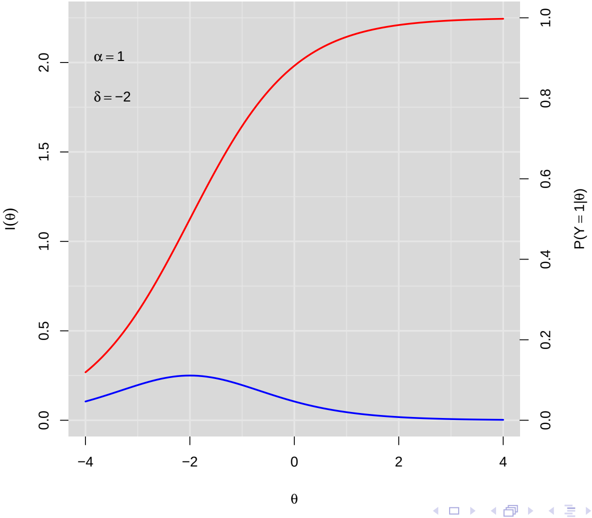
Information

Item information curves



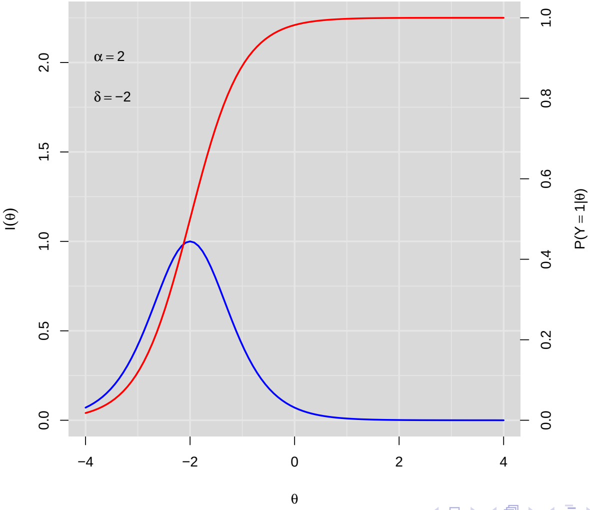
Information

Item information curves



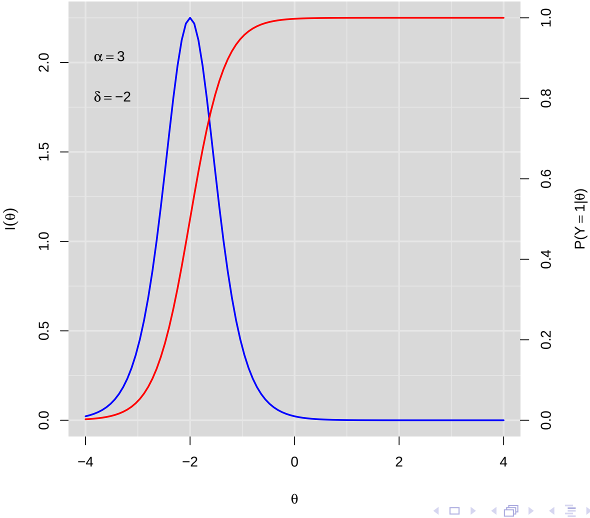
Information

Item information curves



Information

Item information curves



Information

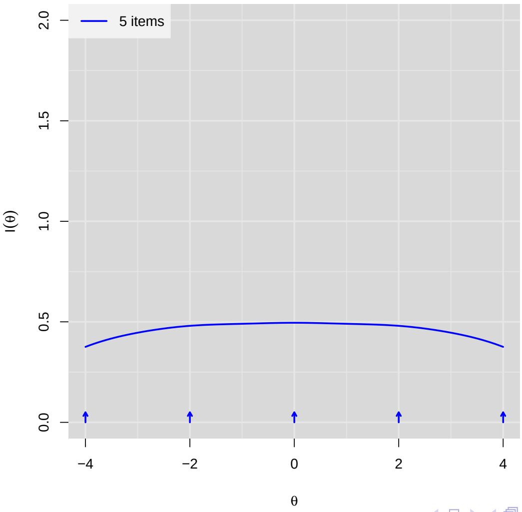
Test information

Information for a test of p items:

$$I(\theta) = \sum_{j=1}^{p} I_j(\theta)$$

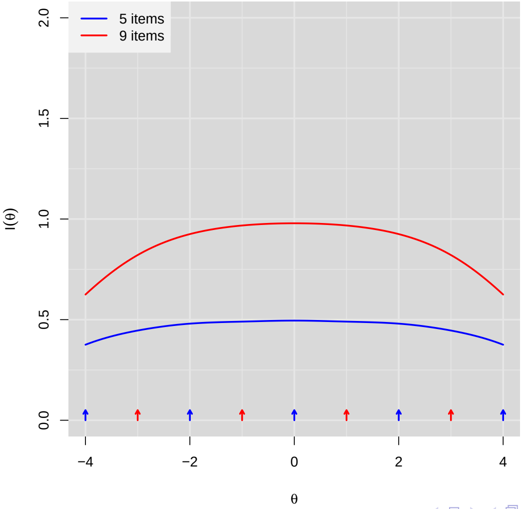
Information

Test information curves



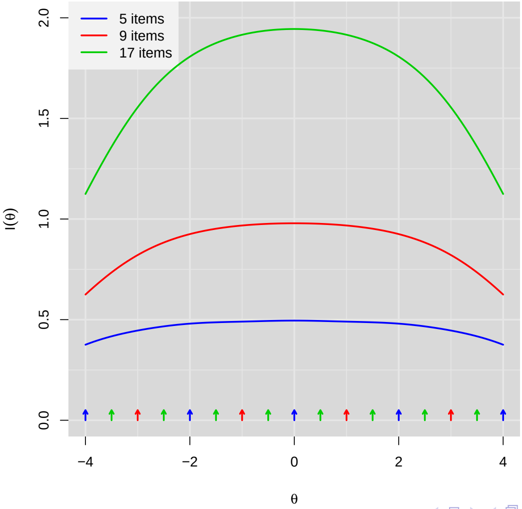
Information

Test information curves



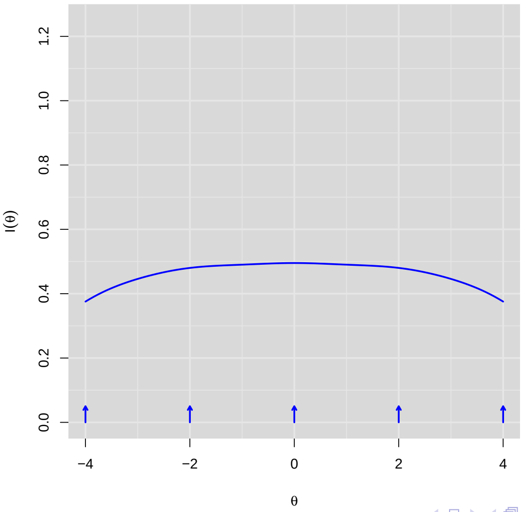
Information

Test information curves



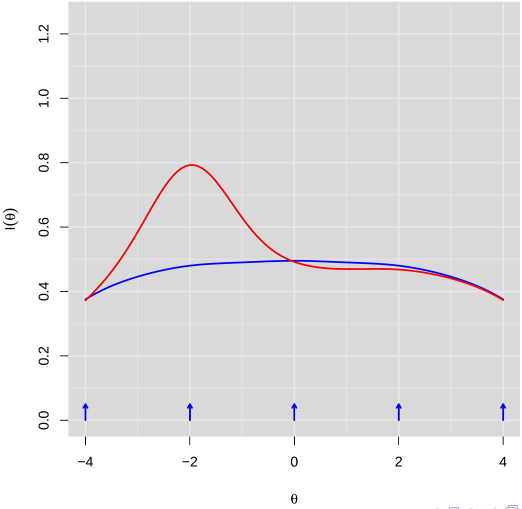
Information

Test information curves



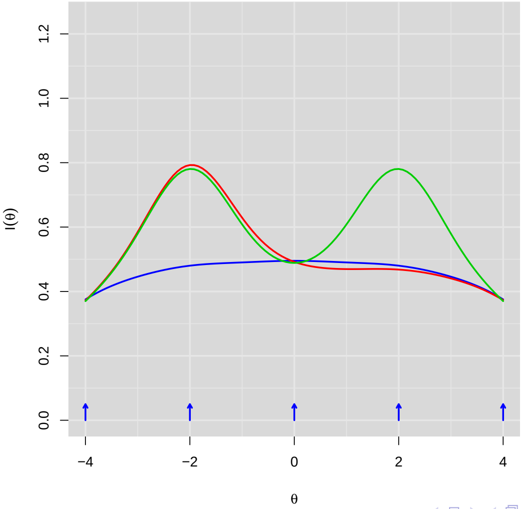
Information

Test information curves



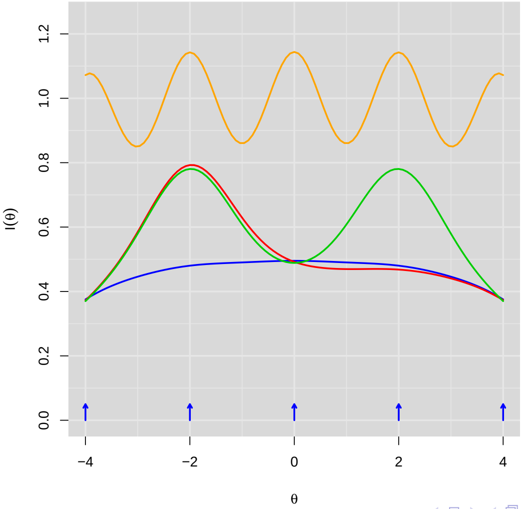
Information

Test information curves



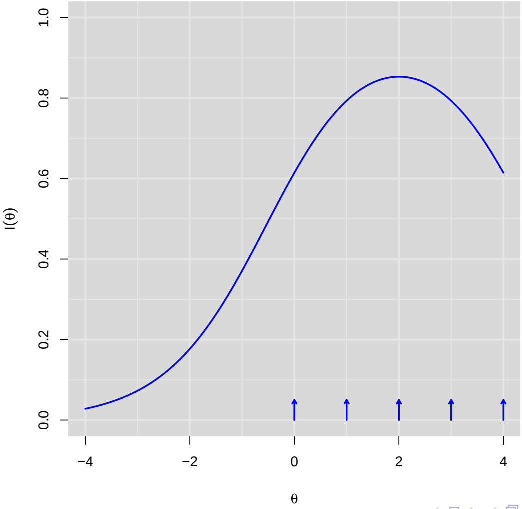
Information

Test information curves



Information

Test information curves



IRT for a sample of *n* individuals

The likelihood

- For the i^{th} individual, we have

IRT for a sample of n individuals

The likelihood

- For the i^{th} individual, we have
  - θ_i, $i = 1, \ldots, n$

IRT for a sample of n individuals

The likelihood

- For the i^{th} individual, we have
  - θ_i, $i = 1, \ldots, n$
  - $(Y_{i1}, \ldots, Y_{ip}) = \mathbf{Y}_i$

IRT for a sample of n individuals

The likelihood

- For the i^{th} individual, we have
  - θ_i, $i = 1, \ldots, n$
  - $(Y_{i1}, \ldots, Y_{ip}) = \mathbf{Y}_i$
  -

$$
\begin{aligned}
P(\mathbf{Y}_i|\theta_i) &= P(Y_{i1} = y_{i1}, \ldots, Y_{ip} = y_{ip}|\theta_i) \\
&= P(Y_{i1} = y_{i1}|\theta_i) \times \cdots \times P(Y_{i1} = y_{ip}|\theta_i)
\end{aligned}
$$

IRT for a sample of n individuals

The likelihood

- For the i^{th} individual, we have
  - θ_i, $i = 1, \ldots, n$
  - $(Y_{i1}, \ldots, Y_{ip}) = \mathbf{Y}_i$
  - $$\begin{aligned} P(\mathbf{Y}_i|\theta_i) &= P(Y_{i1} = y_{i1}, \ldots, Y_{ip} = y_{ip}|\theta_i) \\ &= P(Y_{i1} = y_{i1}|\theta_i) \times \cdots \times P(Y_{i1} = y_{ip}|\theta_i) \end{aligned}$$
- For a sample of n individuals, we have

IRT for a sample of n individuals

The likelihood

- For the i^{th} individual, we have
  - θ_i, $i = 1, \ldots, n$
  - $(Y_{i1}, \ldots, Y_{ip}) = \mathbf{Y}_i$
  -

$$
\begin{aligned}
P(\mathbf{Y}_i | \theta_i) &= P(Y_{i1} = y_{i1}, \ldots, Y_{ip} = y_{ip} | \theta_i) \\
&= P(Y_{i1} = y_{i1} | \theta_i) \times \cdots \times P(Y_{i1} = y_{ip} | \theta_i)
\end{aligned}
$$

- For a sample of n individuals, we have
  - $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$

IRT for a sample of n individuals

The likelihood

- For the i^{th} individual, we have
  - θ_i, $i = 1, \ldots, n$
  - $(Y_{i1}, \ldots, Y_{ip}) = \mathbf{Y}_i$
  - $$\begin{aligned} P(\mathbf{Y}_i|\theta_i) &= P(Y_{i1} = y_{i1}, \ldots, Y_{ip} = y_{ip}|\theta_i) \\ &= P(Y_{i1} = y_{i1}|\theta_i) \times \cdots \times P(Y_{i1} = y_{ip}|\theta_i) \end{aligned}$$

- For a sample of n individuals, we have
  - $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$
  - $$P(\mathbf{Y}_1, \ldots, \mathbf{Y}_n|\theta_1, \ldots, \theta_n) = P(\mathbf{Y}_1|\theta_1) \times \cdots \times P(\mathbf{Y}_n|\theta_n)$$

IRT for a sample of n individuals

The likelihood

- For the i^{th} individual, we have
  - θ_i, $i = 1, \ldots, n$
  - $(Y_{i1}, \ldots, Y_{ip}) = \mathbf{Y}_i$
  - $$\begin{aligned} P(\mathbf{Y}_i|\theta_i) &= P(Y_{i1} = y_{i1}, \ldots, Y_{ip} = y_{ip}|\theta_i) \\ &= P(Y_{i1} = y_{i1}|\theta_i) \times \cdots \times P(Y_{i1} = y_{ip}|\theta_i) \end{aligned}$$

- For a sample of n individuals, we have
  - $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$
  - $$P(\mathbf{Y}_1, \ldots, \mathbf{Y}_n|\theta_1, \ldots, \theta_n) = P(\mathbf{Y}_1|\theta_1) \times \cdots \times P(\mathbf{Y}_n|\theta_n)$$
  - Called the "likelihood."

IRT for a sample of n individuals

Estimating model parameters

- Our model has many parameters: $(\theta_1, \ldots, \theta_n) = \boldsymbol{\theta}$, $(\alpha_1, \ldots, \alpha_p) = \boldsymbol{\alpha}$, and $(\delta_1, \ldots, \delta_p) = \boldsymbol{\delta}$.

IRT for a sample of n individuals

Estimating model parameters

- Our model has many parameters: $(\theta_1, \ldots, \theta_n) = \boldsymbol{\theta}$, $(\alpha_1, \ldots, \alpha_p) = \boldsymbol{\alpha}$, and $(\delta_1, \ldots, \delta_p) = \boldsymbol{\delta}$.
- Likelihood-based estimates: Joint maximum likelihood, marginal maximum likelihood.

IRT for a sample of n individuals

Estimating model parameters

- Our model has many parameters: $(\theta_1, \ldots, \theta_n) = \boldsymbol{\theta}$, $(\alpha_1, \ldots, \alpha_p) = \boldsymbol{\alpha}$, and $(\delta_1, \ldots, \delta_p) = \boldsymbol{\delta}$.
- Likelihood-based estimates: Joint maximum likelihood, marginal maximum likelihood.
- Nonlikelihood-based estimates: Weighted least squares (e.g., in Mplus).

IRT Assumptions

- Unidimensionality.

IRT Assumptions

- Unidimensionality.
  - Example violation: Math word problems

IRT Assumptions

- Unidimensionality.
  - Example violation: Math word problems
- Local independence.

IRT Assumptions

- Unidimensionality.
  - Example violation: Math word problems
- Local independence.
  - Example violation: Testlets

IRT Assumptions

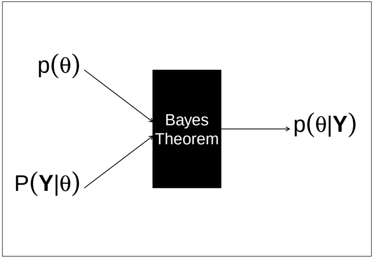
- Unidimensionality.
  - Example violation: Math word problems
- Local independence.
  - Example violation: Testlets
- More sophisticated models are often needed to correct for violations of these assumptions.

Outline

Bayesian inference

Recap

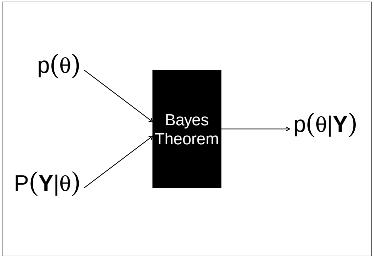
- For Bayesian inference, we need



Bayesian inference

Recap

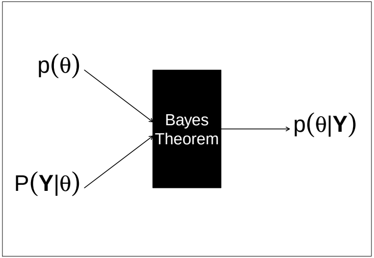
- For Bayesian inference, we need
  1. Likelihood



Bayesian inference

Recap

- For Bayesian inference, we need
  1. Likelihood
  2. Priors for all unknown parameters



Priors for IRT parameters

- $\theta_i \sim \mathrm{N}(0, 1)$

Priors for IRT parameters

- $\theta_i \sim \mathrm{N}(0, 1)$
- $\delta_j \sim \mathrm{N}\left(m_\delta, s_\delta^2\right)$

Priors for IRT parameters

- $\theta_i \sim \mathrm{N}(0, 1)$
- $\delta_j \sim \mathrm{N}\left(m_\delta, s_\delta^2\right)$
- $\alpha_j \sim \mathrm{N}_{(0,\infty)}\left(m_\alpha, s_\delta^2\right)$

Priors for IRT parameters

- $\theta_i \sim \mathrm{N}(0, 1)$
- $\delta_j \sim \mathrm{N}\left(m_\delta, s_\delta^2\right)$
- $\alpha_j \sim \mathrm{N}_{(0,\infty)}\left(m_\alpha, s_\delta^2\right)$
- Values of m_α, s_δ^2, m_δ, s_δ^2 can be chosen reflect prior knowledge of these items (from other studies?).

Priors for IRT parameters

- $\theta_i \sim \mathrm{N}(0, 1)$
- $\delta_j \sim \mathrm{N}\left(m_\delta, s_\delta^2\right)$
- $\alpha_j \sim \mathrm{N}_{(0,\infty)}\left(m_\alpha, s_\delta^2\right)$
- Values of m_α, s_δ^2, m_δ, s_δ^2 can be chosen reflect prior knowledge of these items (from other studies?).
- OR values of s_α^2 and s_δ^2 can be chosen to be large to reflect "ignorance."

Priors for IRT parameters

- $\theta_i \sim \mathrm{N}(0, 1)$
- $\delta_j \sim \mathrm{N}\left(m_\delta, s_\delta^2\right)$
- $\alpha_j \sim \mathrm{N}_{(0,\infty)}\left(m_\alpha, s_\delta^2\right)$
- Values of m_α, s_δ^2, m_δ, s_δ^2 can be chosen reflect prior knowledge of these items (from other studies?).
- OR values of s_α^2 and s_δ^2 can be chosen to be large to reflect "ignorance."
-

$$p(\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\delta}) = p(\theta_1) \cdots p(\theta_n) p(\alpha_1) \cdots p(\alpha_p) p(\delta_1) \cdots p(\delta_p)$$

The posterior distribution

- The posterior distribution for IRT parameters

$$p(\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\delta} | \mathbf{Y}_1, \ldots, \mathbf{Y}_n)$$

The posterior distribution

- The posterior distribution for IRT parameters

$$p(\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\delta} | \mathbf{Y}_1, \ldots, \mathbf{Y}_n)$$

- Too complicated (not a simple Beta(κ_1, κ_2))

The posterior distribution

- The posterior distribution for IRT parameters

$$p(\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\delta} | \mathbf{Y}_1, \ldots, \mathbf{Y}_n)$$

- Too complicated (not a simple Beta(κ_1, κ_2))
- Markov chain Monte Carlo (MCMC) to simulate random draws from the posterior distribution.

The posterior distribution

- The posterior distribution for IRT parameters

$$p(\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\delta} | \mathbf{Y}_1, \ldots, \mathbf{Y}_n)$$

- Too complicated (not a simple Beta(κ_1, κ_2))
- Markov chain Monte Carlo (MCMC) to simulate random draws from the posterior distribution.
- BUGS (WinBUGS, OpenBUGS, JAGS) can do this for you.

The posterior distribution

- The posterior distribution for IRT parameters

$$p(\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\delta} | \mathbf{Y}_1, \ldots, \mathbf{Y}_n)$$

- Too complicated (not a simple Beta(κ_1, κ_2))
- Markov chain Monte Carlo (MCMC) to simulate random draws from the posterior distribution.
- BUGS (WinBUGS, OpenBUGS, JAGS) can do this for you.
  - Open source (free!).

The posterior distribution

- The posterior distribution for IRT parameters

$$p(\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\delta} | \mathbf{Y}_1, \ldots, \mathbf{Y}_n)$$

- Too complicated (not a simple Beta(κ_1, κ_2))
- Markov chain Monte Carlo (MCMC) to simulate random draws from the posterior distribution.
- BUGS (WinBUGS, OpenBUGS, JAGS) can do this for you.
  - Open source (free!).
  - Can be called from other software (R, SAS, Stata).

BUGS code for IRT

Untitled - Notepad

File Edit Format View Help

```
model{
  for (i in 1:n){
    for (j in 1:p){
      Y[i, j] ~ dbern(prob[i, j])
      logit(prob[i, j]) <- alpha[j]*(theta[i] - delta[j])
    }
    theta[i] ~ dnorm(0.0, 1.0)
  }

  for (j in 1:p){
    delta[j] ~ dnorm(m.delta, pr.delta)
    alpha[j] ~ dnorm(m.alpha, pr.alpha) I(0, )
  }
  pr.delta <- pow(s.delta, -2)
  pr.alpha <- pow(s.alpha, -2)
}
```

Outline

Longitudinal Bayesian Item Response Theory

- Easy to change BUGS code to account for longitudinal data.

Longitudinal Bayesian Item Response Theory

- Easy to change BUGS code to account for longitudinal data.
- For examples, see paper "BUGS Code for Item Reponse Theory."

Longitudinal Bayesian Item Response Theory

- Easy to change BUGS code to account for longitudinal data.
- For examples, see paper "BUGS Code for Item Reponse Theory."
- Join Paul's workgroup.