IRT predicted number correct distributions as kernel density estimates

Alexander Weissman
Law School Admission Council, U.S.A.
IRT Workshop, October 2010

Introduction

- Number correct distribution
  - Also known as a “score distribution”
  - Basic outcome of measurement
  - Directly observable quantity
- Item response theory (IRT)
  - Theoretically motivated
  - Assumes underlying latent trait
- What is the connection between the two?
- Why should anyone care?

A debate

- “I’m an IRT person. Why should I worry about observed score distributions if I have item response functions and thetas?”
- “I’m a classical person. Why should I bother about fictitious, complicated, latent trait models when I have the real data?”

Common ground: Smoothing functions

- Assumed in IRT methods
  - Item and test response functions are continuous, differentiable
  - Smooth distributions follow naturally from model
- Needed in classical (observed score) methods
  - Equipercentile equating
  - Smoothing needed to deal with discreteness of data
  - Smooth distributions not intrinsic

Smoothing philosophies

Smooth functions are needed

Smooth functions are assumed

Observed Score Tradition

True Score Tradition

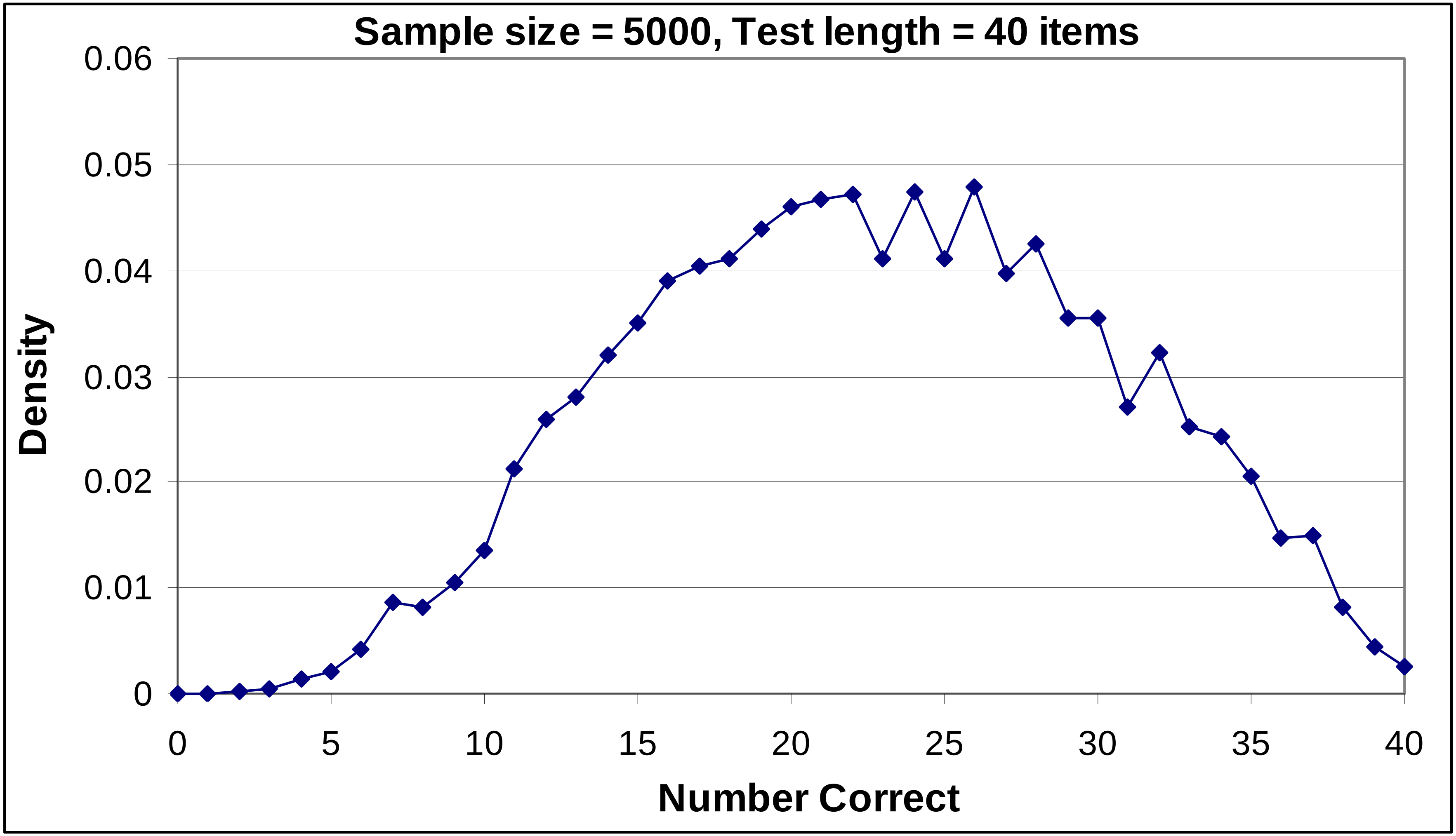
Interpretation:
Repeated sampling

Interpretation:
Underlying function

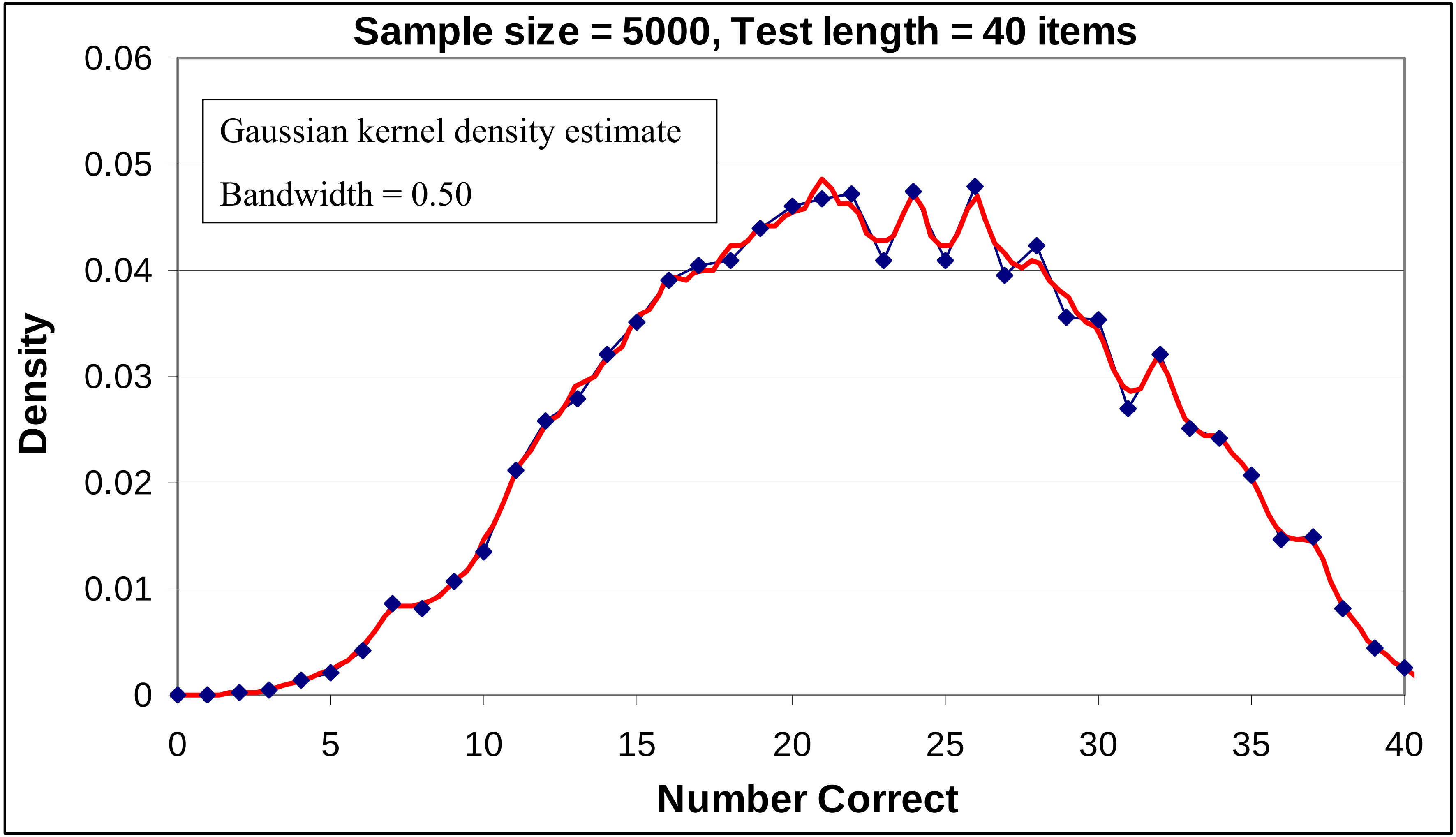
Smoothing methodologies

- Observed score tradition
  - Log linear smoothing
    - Polynomial degree / moments
  - Splines
    - Number and placement of knots
    - Polynomial degree, continuity rules
  - Kernel density estimation
    - Kernel function
    - Bandwidth
- True score tradition
  - IRT
    - Marginal maximum likelihood

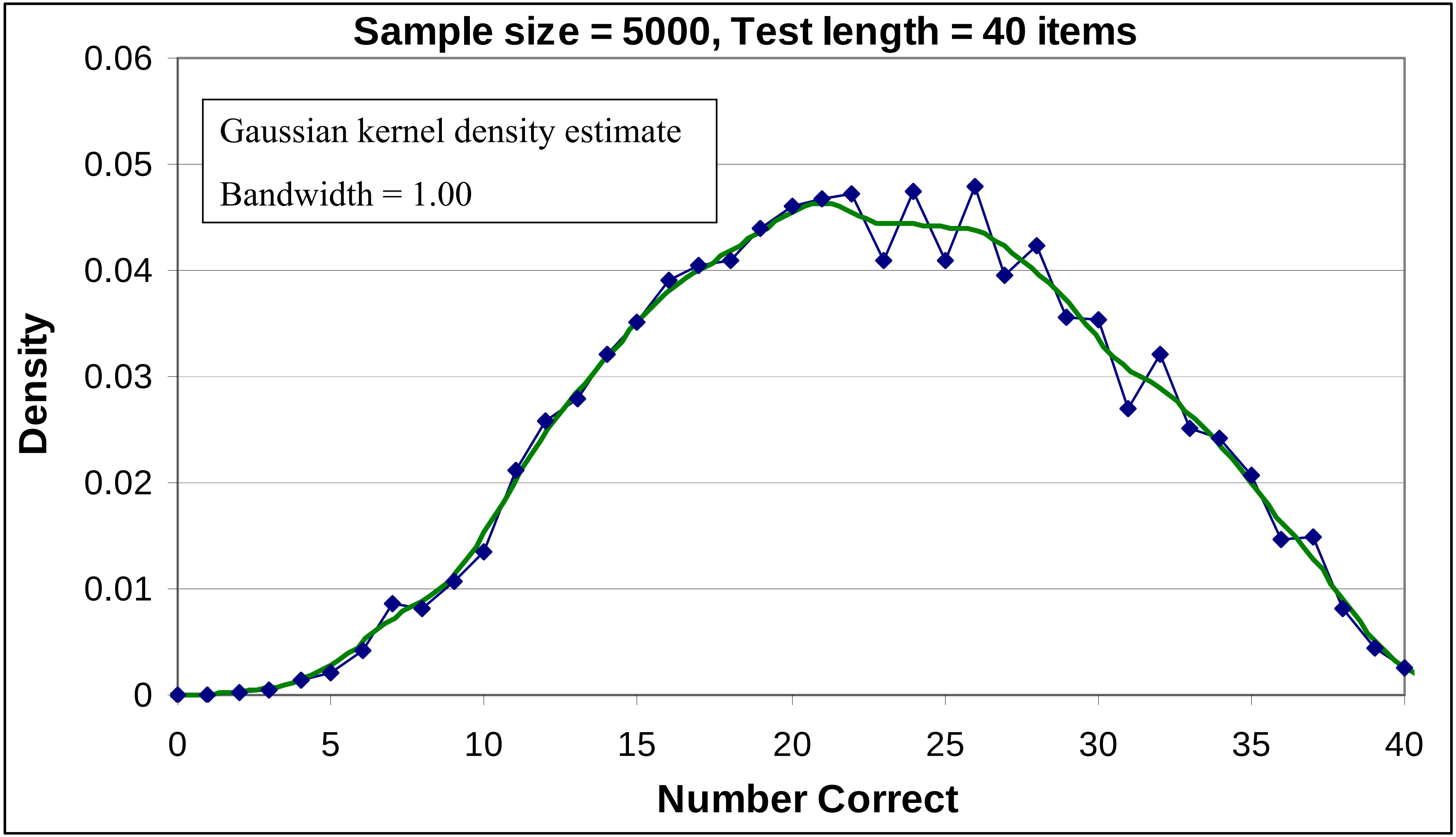
Empirical NC distribution - example



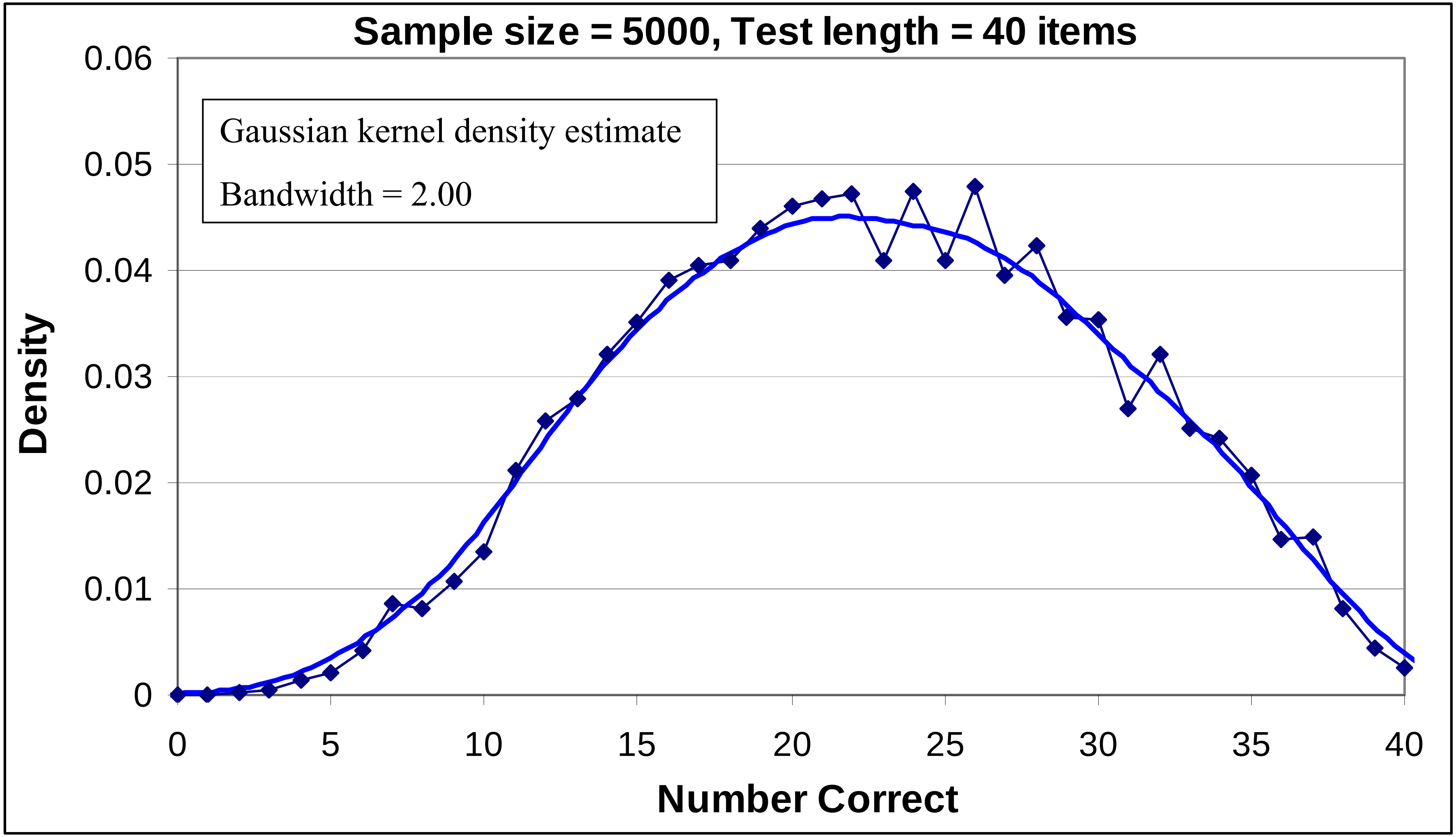
Smooth distribution



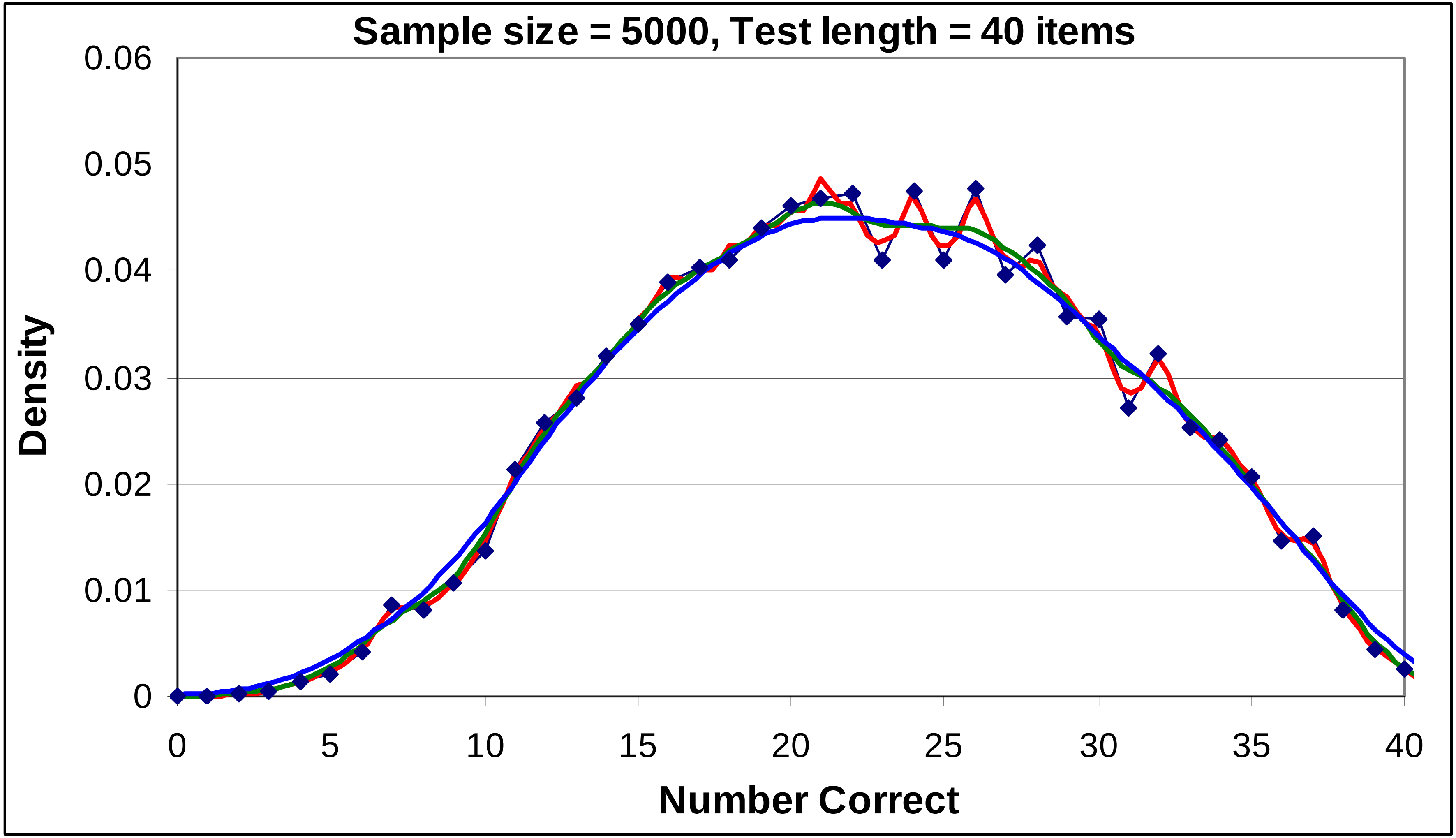
Another smooth distribution



Yet another smooth distribution



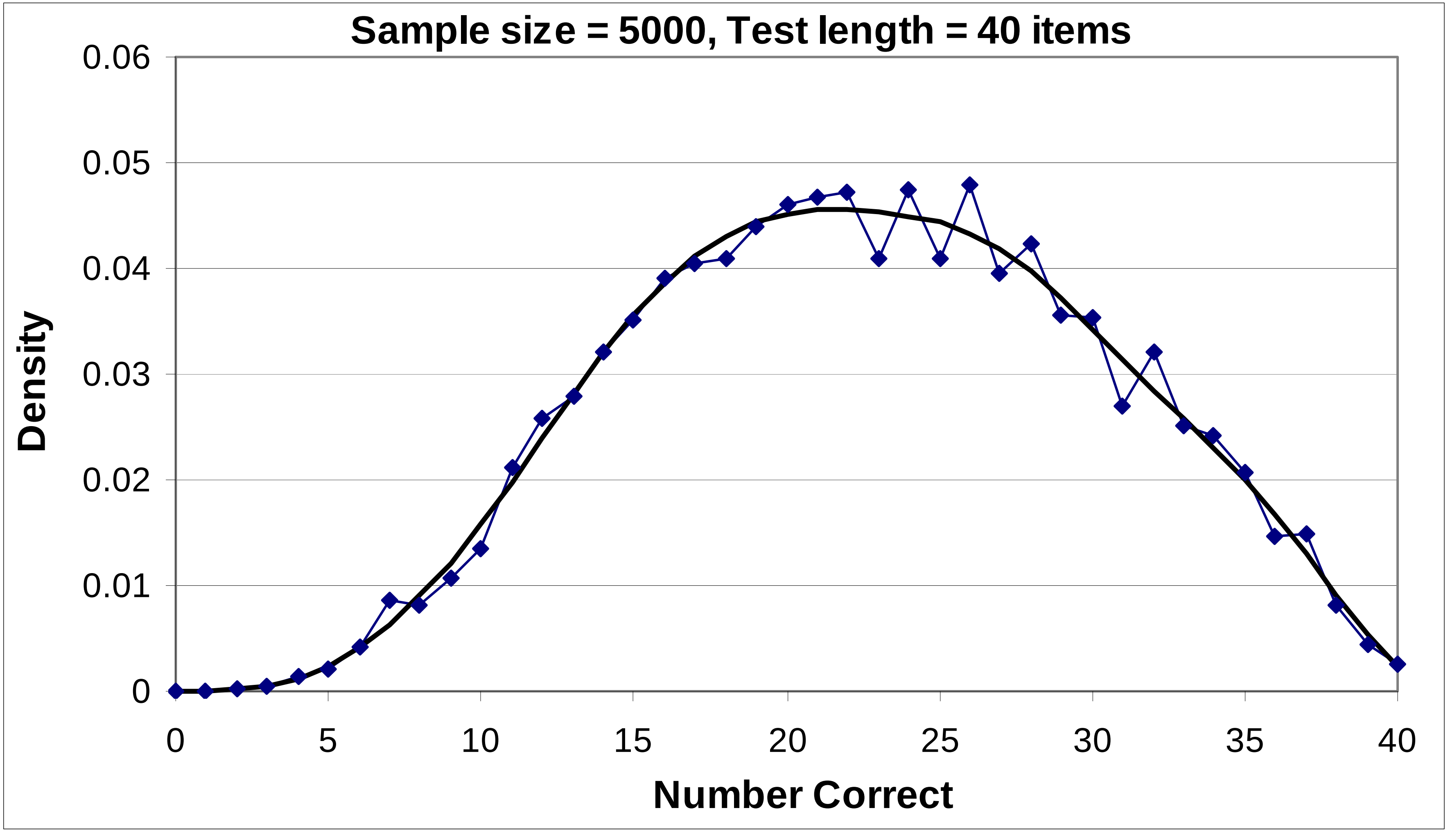
Which one? Or should we try another?



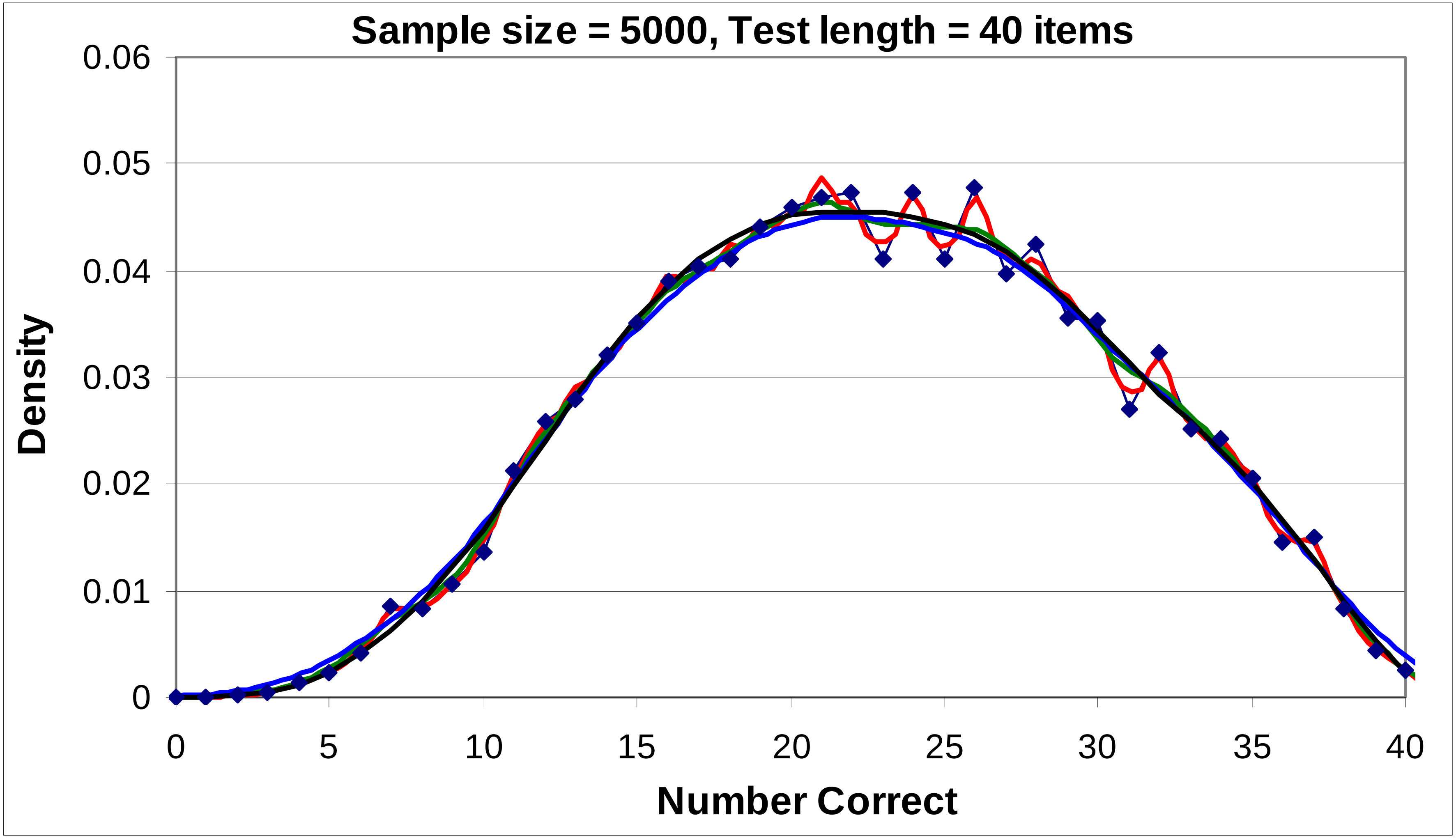
IRT predicted number correct distribution

- Suppose instead we use IRT
  - Estimate item parameters using marginal maximum likelihood
  - Obtain the predicted number correct distribution

IRT predicted NC distribution



Comparison with observed score KDEs



Obtaining predicted NC distributions

- Given
  - A set of J dichotomously-scored items,
  - Item parameter vector $\boldsymbol{\omega}_j$ for each item j,
  - A specific theta value θ_k
- Find
  - The predicted number correct distribution $f(r|\theta_k)$, $r = 0, 1, \ldots, J$.

Lord-Wingersky recursion

- See Lord & Wingersky (1984, *Applied Psychological Measurement*); Kolen & Brennan (1995)
- Let $p_{jk} = P\left(X_j = 1 \middle| \theta_k, \boldsymbol{\omega}_j\right)$

  $q_{jk} = 1 - p_{jk}$

- Start with $j = 1$. Then $f\left(r = 0 \middle| \theta_k\right) = q_{jk}$

  $f\left(r = 1 \middle| \theta_k\right) = p_{jk}$

Lord-Wingersky recursion - continued

- Given $j > 1$ items and $f(r|\theta_k)$ for $r < j$

NC, $r < j$	$f(r\|\theta_k)$	p_{jk}	q_{jk}
0	$f(0\|\theta_k)$	$f(0\|\theta_k) \cdot p_{jk}$	$f(0\|\theta_k) \cdot q_{jk}$
1	$f(1\|\theta_k)$	$f(1\|\theta_k) \cdot p_{jk}$	$f(1\|\theta_k) \cdot q_{jk}$
…			
j-1	$f(j-1\|\theta_k)$	$f(j-1\|\theta_k) \cdot p_{jk}$	$f(j-1\|\theta_k) \cdot q_{jk}$

Lord-Wingersky recursion - continued

- Now collect terms to obtain $f^*(r|\theta_k)$ for $r \leq j$

NC, $r < j$	$f(r\|\theta_k)$	p_{jk}	q_{jk}
0	$f(0\|\theta_k)$	$f(0\|\theta_k) \bullet p_{jk}$	$f(0\|\theta_k) \bullet q_{jk}$
1	$f(1\|\theta_k)$	$f(1\|\theta_k) \bullet p_{jk}$	$f(1\|\theta_k) \bullet q_{jk}$
…			
j-1	$f(j\text{-}1\|\theta_k)$	$f(j\text{-}1\|\theta_k) \bullet p_{jk}$	$f(j\text{-}1\|\theta_k) \bullet q_{jk}$

$$f^*(0|\theta_k)$$

$$f^*(1|\theta_k)$$

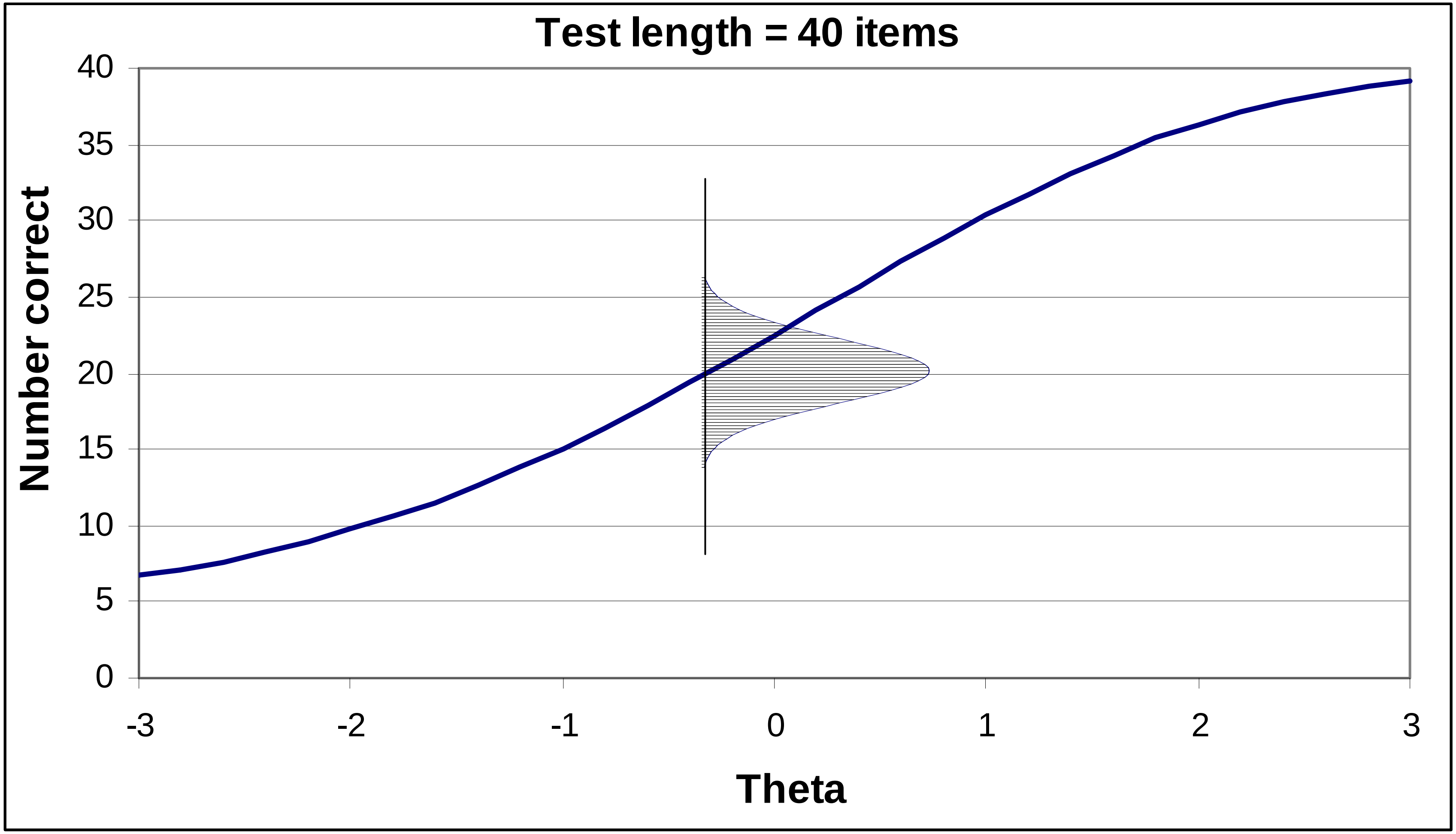
$$f^*(j\text{-}1|\theta_k)$$

$$f^*(j|\theta_k)$$

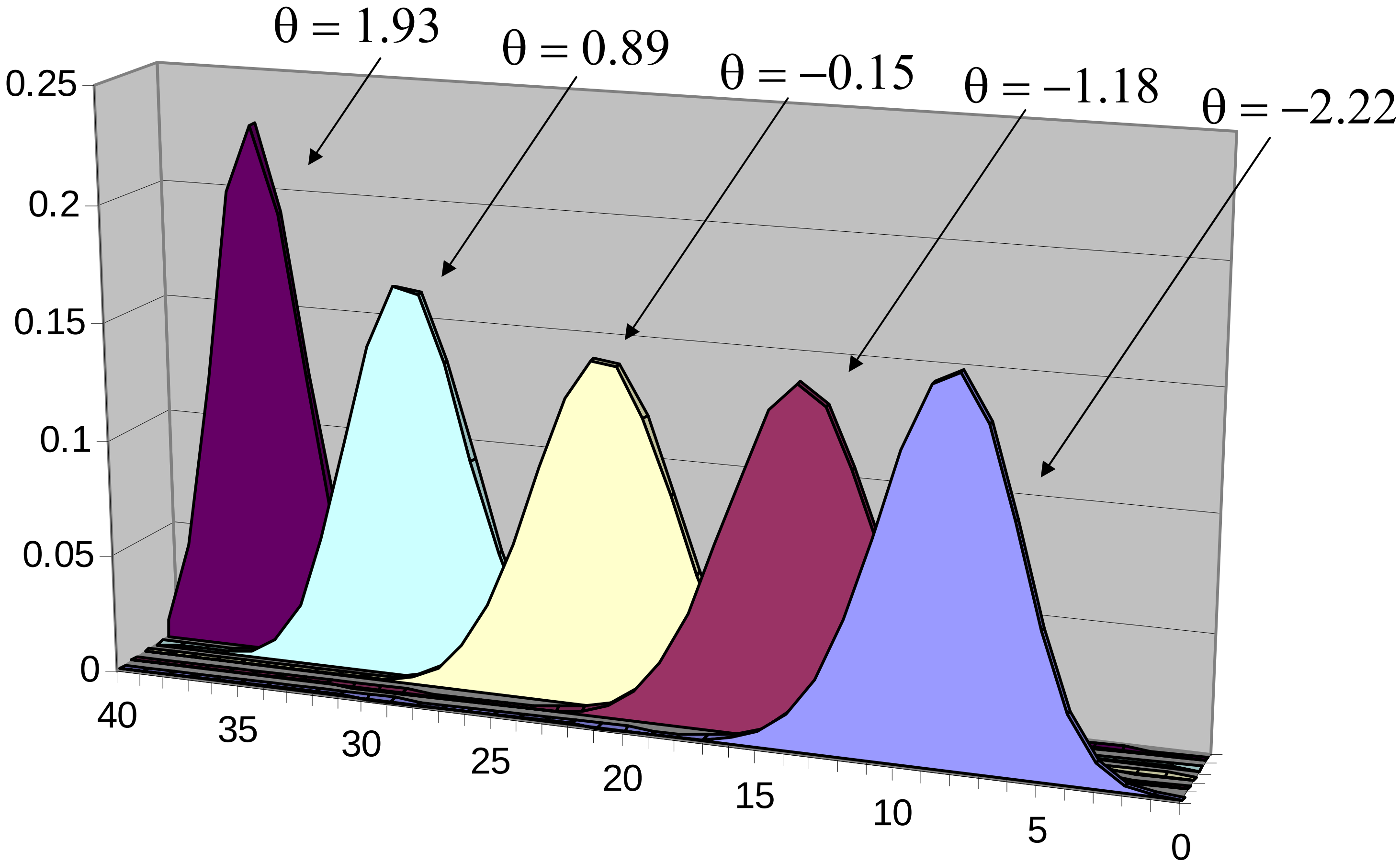
An IRT kernel

- What we end up doing is constructing a kernel density estimate
- In addition to satisfying the usual properties of a kernel, it has a number of advantages
  - No need to choose functional form
  - No need to select a bandwidth
  - Interpretable as a conditional distribution
  - Conditional expectation is the test response function

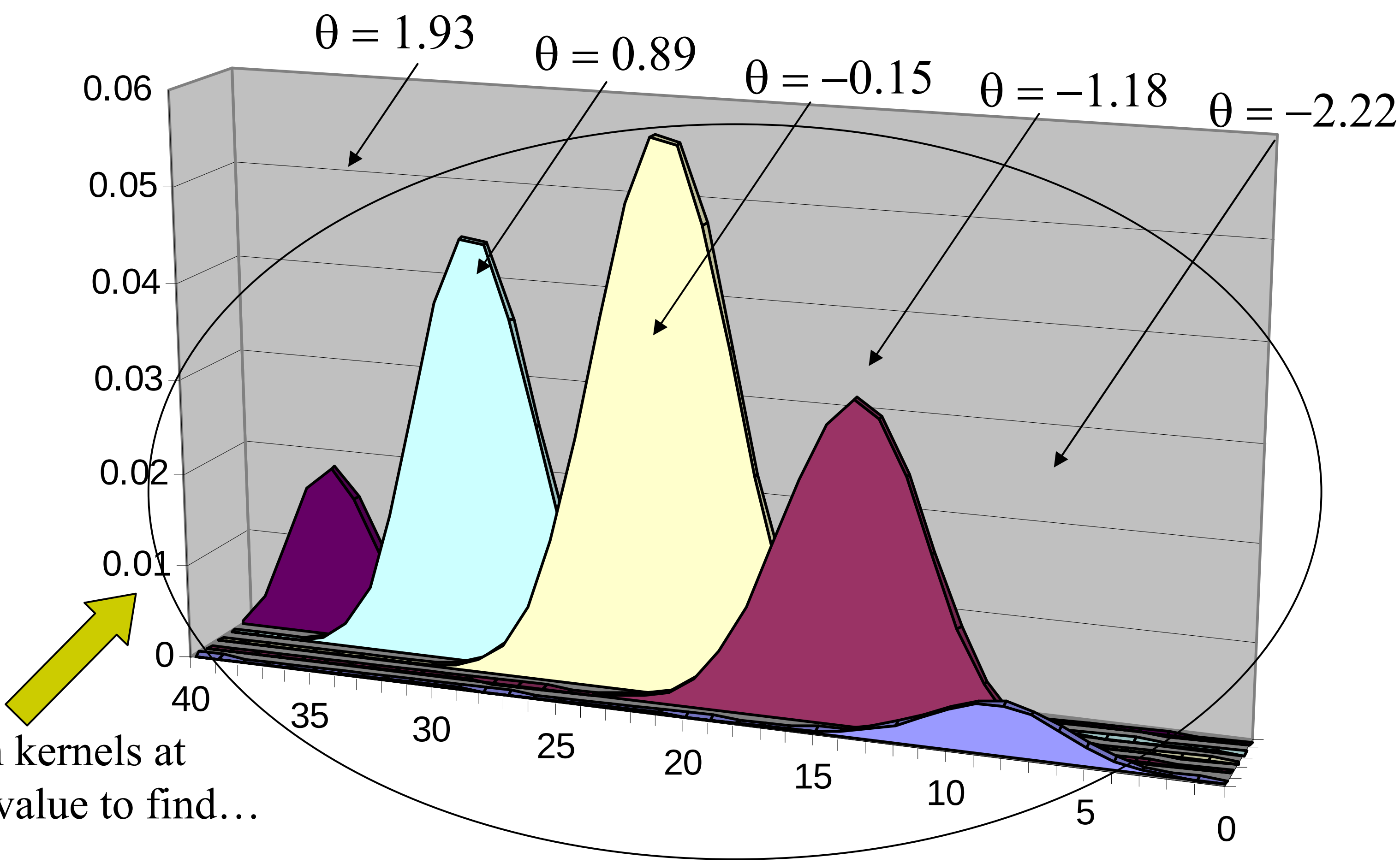
Test response function



IRT kernels - conditionals

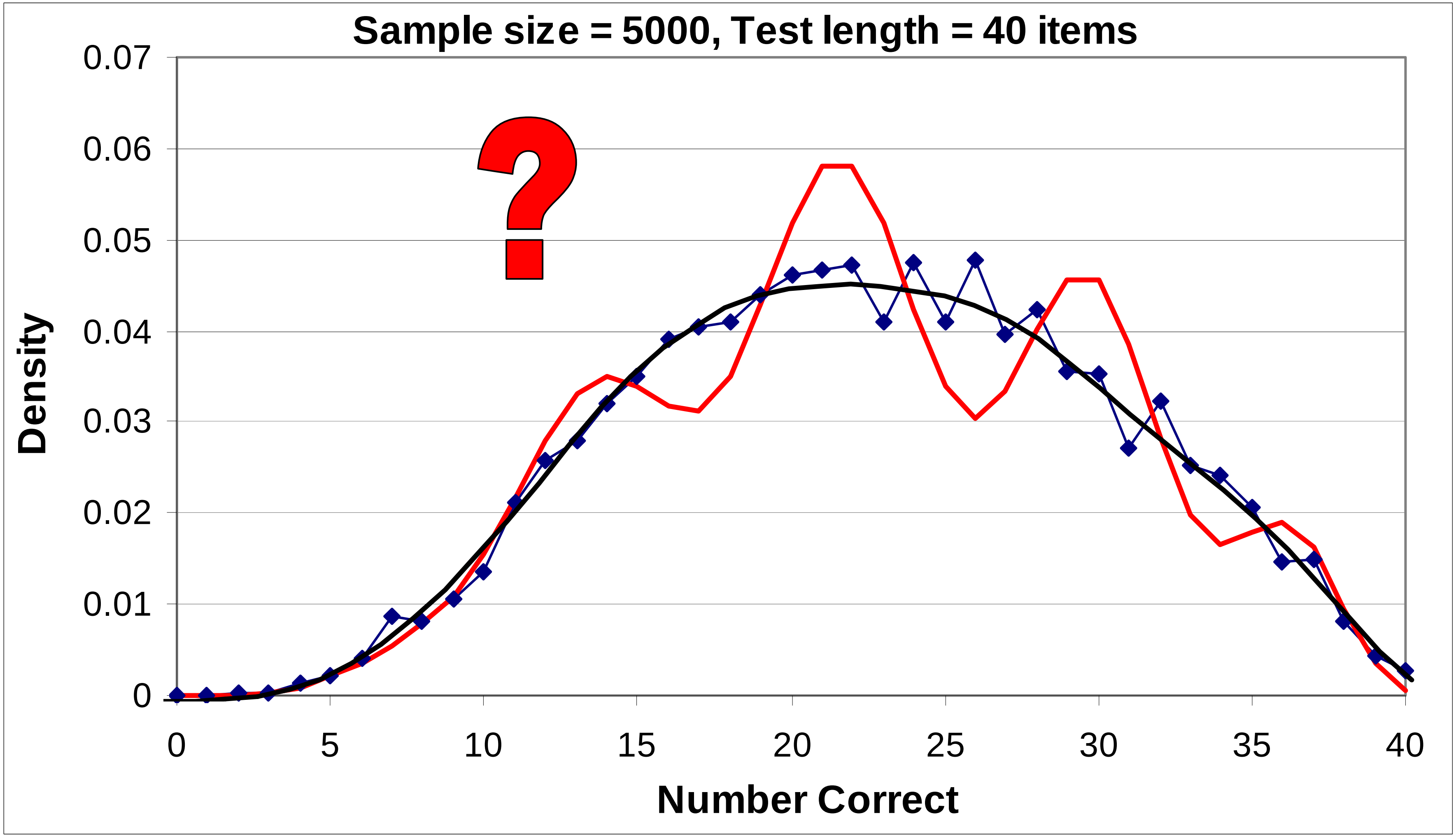


IRT kernels – weighted by posterior



Then sum kernels at each NC value to find…

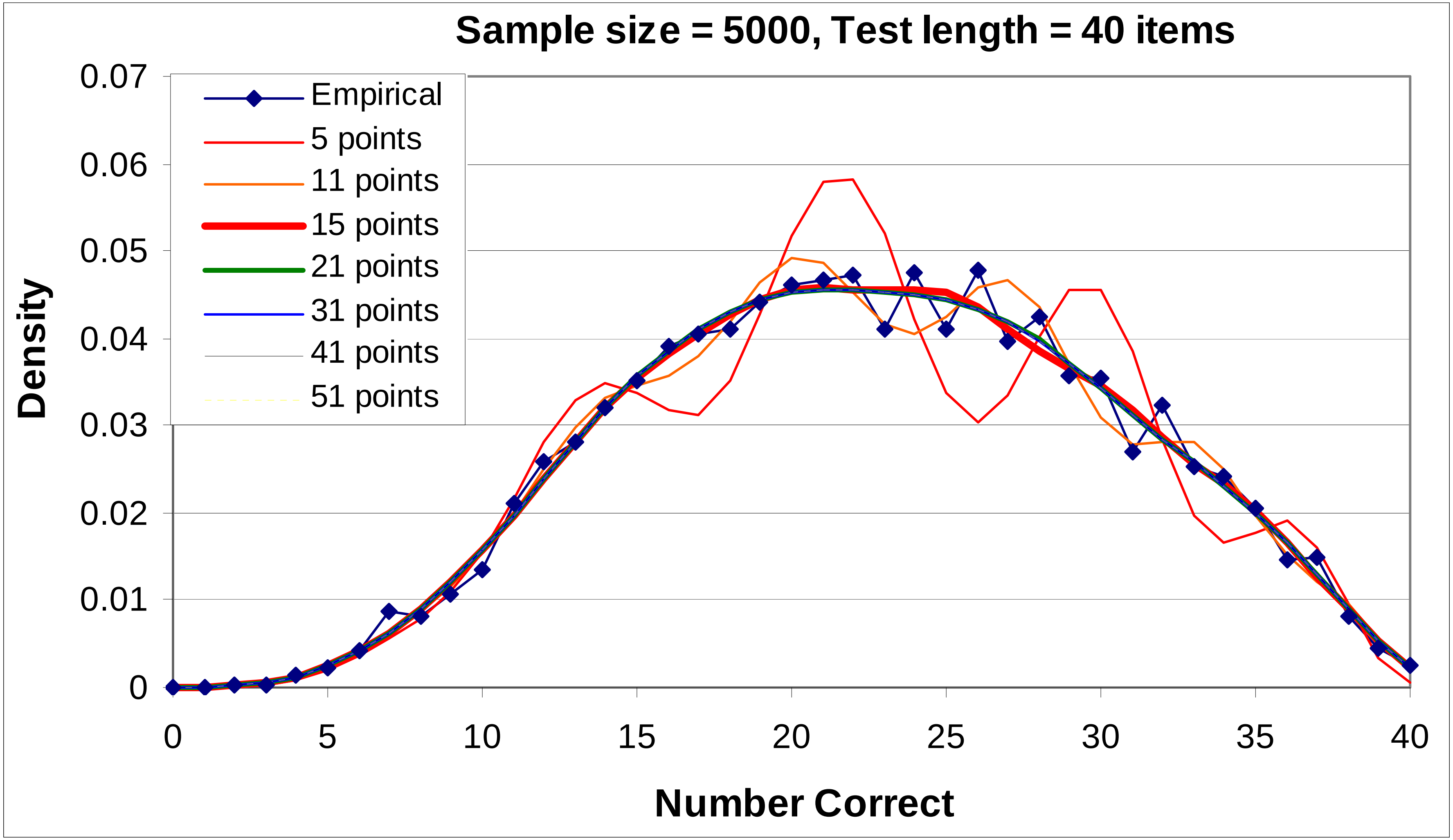
IRT kernel density estimate: Predicted NC distribution



What happened?

- The number of quadrature points used in the MML estimation makes a difference!
  - As the number increases, the predicted NC distribution becomes smoother
  - This smoothness reaches a limit

Different numbers of quadrature points



Assessing MML estimation

- Number of quadrature points
  - “Rules of thumb” appear to be supported
  - We can determine when we have enough points
- Model-data fit
  - Predicted versus observed NC correct distribution provides another criterion for goodness-of-fit

Revisiting some questions

- "I'm an IRT person. Why should I worry about observed score distributions if I have item response functions and thetas?"
  - IRT predicted number correct distributions should match closely with the observed score distributions
  - For a given dataset, there is a threshold for the number of quadrature points needed in MML estimation
  - Smoothing is a natural outcome of model

Revisiting some questions

- "I'm a classical person. Why should I bother about fictitious, complicated, latent trait models when I have the real data?"
  - Less arbitrary smoothing methodology, one instead of many disparate choices
  - Results on more solid theoretical ground
  - IRT parameter estimates are available, though they need not be used

End of presentation

For more information, contact:

Alexander Weissman

Law School Admission Council, U.S.A.

aweissman@lsac.org