Inference, robustness, and essential unidimensionality

Brian W. Junker
Dept of Statistics
Carnegie Mellon University
Pittsburgh PA 15213
(412) 268-8873
brian@stat.cmu.edu

Thanks to Stout and the UIUC workgroup, and especially to Douglas and Kim for letting me repeat their results here.

Inference, robustness, and essential unidimensionality

- Notation and basic concepts

- $d_E = 1$ and useful models

- Robustness neighborhoods of $d_L = 1$

- Neighborhood measures

- Inference under $d_E = 1$

- What are we estimating when $d_E > 1$?

Notation and basic concepts

- J items (later: $J \to \infty$): $\underline{X}_J = (X_1, X_2, \ldots, X_J)$

$$X_j = 0 \quad \text{or} \quad 1$$

- Summary scores:

$$\overline{X}_J = \frac{1}{J}\sum_{j=1}^{J} X_j \quad \text{or} \quad \overline{A}_J = \frac{1}{J}\sum_{j=1}^{J} a_j X_j$$

where $a_j \le M < \infty$ and $0 < \epsilon \le \frac{1}{J}\sum_j a_j$.

- Expected ("true") scores:

$$\begin{aligned} P_j(\theta) &= E[X_j|\theta] \\ \overline{P}_J(\theta) &= E[\overline{X}_J|\theta] \end{aligned}$$

$$\begin{aligned} A_j(\theta) &= E[a_j X_j|\theta] \\ \overline{A}_J(\theta) &= E[\overline{A}_J|\theta] \end{aligned}$$

$\underline{d_L = 1}$ **(Strictly Unidimensional; MH)**

- LI: $(X_1, \ldots, X_J)$ independent, given Θ:

$$P[\underline{X}_J = \underline{x}_J] = \int \prod_{j=1}^{J} P_j(\theta)^{x_j}(1-P_j(\theta))^{1-x_j}\, dF(\theta)$$

- M: $P_j(\theta)$ nondecr. in θ.

  Equivalently:

  * $P[X_j > c|\theta]$ nondecr in θ; or

  * $A_j(\theta)$ nondecr. in θ $\forall$ $\{a_j\}$: $a_j > 0$.

- $d_L = 1$: The above can be done for a unidimensional θ.

$d_E = 1$ (Essentially Unidimensional)

• EI (essentially independent):

$$\lim_{J\to\infty} \binom{J}{2}^{-1} \sum\sum_{1\le i<j\le J} |\mathrm{Cov}\,(X_i, X_j|\theta)| = 0$$

• LAD (locally asymptotically discriminating):

At each θ, $\overline{P}_J(\theta)$ has a positive slope (discrimination) bounded away from 0 as $J \to \infty$.

(If $\overline{P}'_J(\theta)$ exists, then $\overline{P}'_J(\theta) \ge \varepsilon_\theta$, as $J \to \infty$.)

• $d_E = 1$: The above can be done for a unidimensional θ.

$d_E = 1$, $d_L = 1$, and Useful Models

- EI implies

$$\overline{A}_J \text{ consistently estimates } \overline{A}_J(\theta)$$

and $d_E = 1$ [EI + LAD] implies

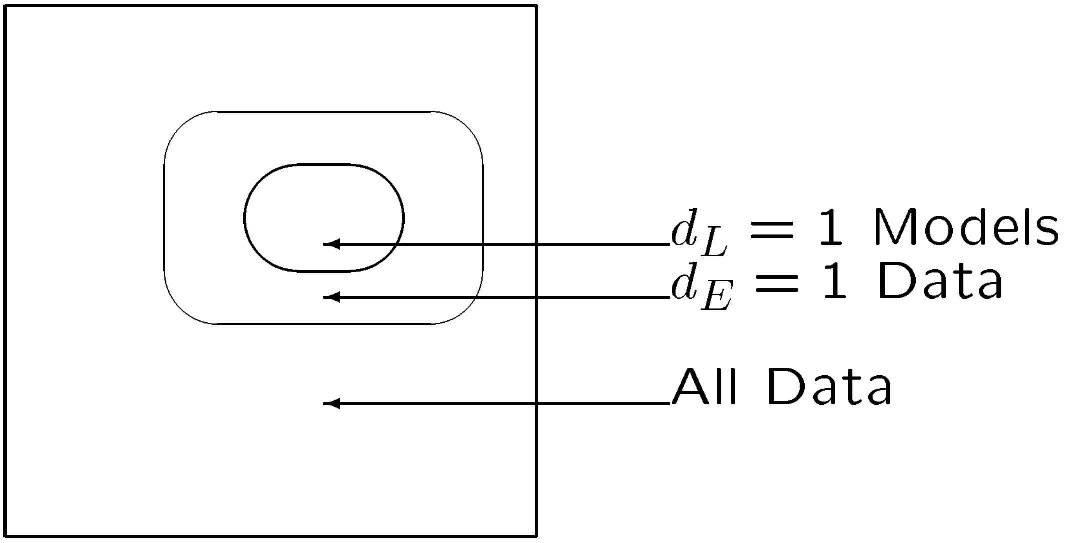
$$\overline{A}_J^{-1}(\overline{A}_J) \text{ consistently estimates } \theta$$

for all bounded scores $|a_j| \leq M < \infty$.

- $(d_L = 1)$ + (LAD) $\Rightarrow d_E = 1$
- θ is unique up to monotone transformation (*ordinal uniqueness*).
- Polytomous definitions and extensions: Junker (1991, *Psy*).

- *Useful* models (Junker, 1993 *Ann Stat*):
  * consistent estimation of θ
  * $\overline{P}_J(\theta)$ monotone (Stout's *weak monotonicity*, WM)
  * Different subjects have different θ's

Robustness neighborhoods of $d_L = 1$



- Drasgow and Parsons (1983 *APM*):

> "...it seems clear that researchers should be more concerned with the robustness of estimation techniques to minor violations of dimensionality assumptions than with the possibly neverending task of measuring all latent variables that underlie responses in a particular content domain."

- $d_E = 1$ identifies a *dominant* trait θ that can still be consistently estimated (there is a signal to detect!) amid possibly many minor traits.

- Robustness idea: any method that we would like to apply in $d_L = 1$ should be tried out on the testbed of $d_E = 1$.

Neighborhood measures

• The study of $d_E = 1$ is the study of robustness of ability estimators to variations from a strict $d_L = 1$ latent structure.

• LAD is an identifiability condition: *if* a dominant θ exists, *then* it must be monotonically related to $\overline{X}_J$ (or $\overline{A}_J$).

$\rightarrow$ Are we in $d_E = 1$ setting?

$\rightarrow$ How "far" from $d_L = 1$ are we?

Neighborhood measures:

$d_E = 1$**?**

- X_1, X_2, ... <u>dichotomous</u>.
- **Test for** $d_E = 1$ (Stout, 1987 *Psy*)

1. AT1 item cluster—as unidimensional as possible.
2. AT2 item cluster—like PT dimensionally, but like AT1 in difficulty.
3. PT item cluster—control for ability with X_+ on PT.
4. Compare $d = 1$ variance-estimate vs. naive variance-estimate in AT1, using AT2 for bias correction, averaged over X_+ on PT.

- Although $d_E = 1$ is not *defined* in terms of $\underline{X}_J$ alone, it is *testable* from $\underline{X}_J$.
- Stout (1990, *Psy*), Nandakumar & Stout (1993, *JES*), Nandakumar (1991, *JEM*).

Neighborhood measures:
How far from $d_L = 1$?

Junker and Stout (1991, *EMPS*):

- $d_E = 1$ implies:

$$\epsilon_J = \binom{J}{2}^{-1} \sum\sum_{1 \le i < j \le J} \left[\int |\mathrm{Cov}\,(X_i, X_j|\theta)| f(\theta) d\theta \right] \to 0$$

as $J \to \infty$.

Can estimate directly with

$$\hat{\epsilon}^0_J = \binom{J}{2}^{-1} \sum\sum_{1 \le i < j \le J} \left[\sum_{k=0}^{J} |\widehat{\mathrm{Cov}}\,(X_i, X_j|X_+ = k)| \, \frac{N_k}{N} \right]$$

- Total score $X_+ = \sum_j X_j$

- N_k examinees score k out of J

- $\widehat{\mathrm{Cov}}\,(X_i, X_j|X_+ = k) =$
  $(1/N_k) \sum_{n=1}^{N_k} (x_{ni} - \overline{x}_i)(x_{nj} - \overline{x}_j)$

Neighborhood measures:
How far from $d_L = 1$?

Ideally,

$$\begin{aligned} \hat{\epsilon}^0_J &\approx 0 \text{ if } d_E = 1; \\ \hat{\epsilon}^0_J &\gg 0 \text{ if } d_E > 1. \end{aligned}$$

But often $\hat{\epsilon}^0_J$ too big for $d_E = 1$.

Reasons:

• Nonzero covariances due to natural random variability in the data;

• $\mathrm{Cov}\,(X_i, X_j | X_+ = k) < 0$ may occur *even though* $\mathrm{Cov}\,(X_i, X_j | \Theta = \theta) \equiv 0 \ \forall \ \theta$ (Junker, 1993 *Ann Stat*).

$|\widehat{\mathrm{Cov}}\,(X_i, X_j | X_+ = k)|$ counts these *against* $d = 1$ when they should count *for* $d = 1$!

Neighborhood measures:
How far from $d_L = 1$?

- Fix bias in $\hat{\epsilon}^0_J$:

1. $\{\lambda_{j2} : \ j = 1, \ldots, J\}$ = second factor loadings of principal-factors factor analysis.

2. Omit $\widehat{\text{Cov}}\,(X_i, X_j | X_+ = k)$ whenever:
   * $|\lambda_{i2}|$ or $|\lambda_{j2}| \leq$ fixed cutoff M; or
   * $\text{sign}\{\widehat{\text{Cov}}\,(X_i, X_j | X_+ = k)\} \neq \text{sign}\{\lambda_{i2}\lambda_{j2}\}$.

- $\boxed{d = 1\text{: } \theta_1 = \text{true; } \theta_2 = \text{noise}}$
  * $\lambda_{2j} \approx 0$; $\text{Cov}\,(X_i, X_j | X_+) < 0$ expected but "half" the $\lambda_{i2} \cdot \lambda_{j2} > 0$.

- $\boxed{d = 2\text{, simple structure: } \theta_1 = \text{“g”}\,;\ \theta_2 = \text{“bipolar”}}$
  * $\lambda_{i2} \cdot \lambda_{j2} > 0 \Rightarrow X_i,\ X_j$ probably same subtest $\Rightarrow \text{Cov}\,(X_i, X_j | X_+) > 0$;
  * $\lambda_{i2} \cdot \lambda_{j2} < 0 \Rightarrow X_i,\ X_j$ probably different subtests $\Rightarrow \text{Cov}\,(X_i, X_j | X_+) < 0$

Neighborhood measures:
How far from $d_L = 1$?

$J = 40,\ N = 2000$	T^*	$100\hat{\epsilon}_J$				$100\epsilon_J$	d_L	Ref.†
M (tuning parameter)		.00	.10	.15	.20			Comp.
2PL ($d_L = 1$)	−2.24	.46	.07	.07	.00	0.00	1	
Simple MD2PL	5.40	2.32	2.32	2.32	2.32	2.14	2	38°
3PL ($d_L = 1$)	−0.39	.71	.24	.15	.05	0.00	1	
Simple MD3PL	8.23	1.83	1.78	1.78	1.73	1.57	2	49°
Simple MD2PL, $\rho = 0.0$	18.19	3.73	3.73	3.73	3.73	3.43	2	27°
Simple MD2PL, $\rho = 0.3$	13.41	2.65	2.65	2.65	2.65	2.67	2	44°
Simple MD2PL, $\rho = 0.5$	10.47	2.06	2.06	2.06	2.06	1.94	2	45°
Simple MD2PL, $\rho = 0.7$	8.45	1.41	1.41	1.41	1.41	1.19	2	45°
Simple MD2PL, $\rho = 0.9$	−0.35	.79	.55	.11	.02	0.41	2	45°

* Test statistic from Stout (1987). † See Wang (1986).

Simple structure results for ϵ_J and $\hat{\epsilon}_J$.

$J = 40,\ N = 2000$	T^*	$100\hat{\epsilon}_J$				$100\epsilon_J$	d_L	Ref.†
M (tuning parameter)		.00	.10	.15	.20			Comp.
2PL ($d_L = 1$)	−2.24	.46	.07	.07	.00	0.00	1	
Mixed MD2PL	4.03	.87	.60	.35	.16	0.31	2	44°
Mixed MD2PL, $\rho = 0.0$	6.91	.84	.49	.39	.31	0.44	2	50°
Mixed MD2PL, $\rho = 0.3$	3.58	.74	.34	.23	.16	0.31	2	46°
Mixed MD2PL, $\rho = 0.5$	−3.64	.62	.27	.11	.01	0.22	2	45°
Mixed MD2PL, $\rho = 0.7$	1.03	.58	.10	.01	.00	0.14	2	45°
Mixed MD2PL, $\rho = 0.9$	−1.42	.53	.03	.00	.00	0.05	2	45°
Paragr. Comp. ($\rho = 0.0$)	2.32	.63	.08	.02	.00	0.05	9	6°

* Test statistic from Stout (1987). † See Wang (1986).

Mixed structure results for ϵ_J and $\hat{\epsilon}_J$.

Neighborhood measures: DETECT

H.-R. Kim (1994 *dissertation*)

- Replace

$$\sum_{k=0}^{J} |\widehat{\mathsf{Cov}}\,(X_i, X_j | X_+ = k)| \, \frac{N_k}{N}$$

with

$$\widehat{\mathsf{Cov}}_{ij} = \sum_{k=0}^{J-2} \widehat{\mathsf{Cov}}\,(X_i, X_j | S_{ij} = k) \, \frac{N_k}{N}$$

where $S_{ij} = X_+ - X_i - X_j$: rest-score!

- Form "preliminary item clusters"
- New statistic:

$$DETECT = \binom{J}{2}^{-1} \sum\sum_{1 \le i < j \le J} \delta_{ij} \left[\widehat{\mathsf{Cov}}_{ij} - \overline{\mathsf{Cov}} \right]$$

where

$$\overline{\mathsf{Cov}} = \binom{J}{2}^{-1} \sum\sum_{1 \le i < j \le J} \widehat{\mathsf{Cov}}_{ij}$$

and

$$\delta_{ij} = \begin{cases} +1, & \text{if } X_i \text{ and } X_j \text{ are in the same cluster;} \\ -1, & \text{if not.} \end{cases}$$

Neighborhood measures: DETECT

- δ_{ij} gives a rule for weighting evidence from item pairs from same/different clusters (like "throwing out covariances" in $\widehat{\epsilon}_J$).

- DETECT $\leq$

$$DETECT_{max} = \binom{J}{2}^{-1} \sum\sum_{1\leq i<j\leq J} |\widehat{\mathsf{Cov}}_{ij} - \overline{\mathsf{Cov}}|$$

- Multiple item clusters extends methodology beyond 2-factor / 2-cluster method of $\widehat{\epsilon}_J$.

- Preliminary clusters: expert opinion, F.A., hierarchical cluster analysis (HCA), exhaustive search...

- Combine with HCA, etc., as exploratory device to find "the right number of clusters".

Neighborhood measures: DETECT

KIM'S DETECT PIC, P. 37

Source: Kim (1994)

Neighborhood measures: DETECT

KIM'S DETECT PIC, P. 36

Source: Kim (1994)

Neighborhood measures: DETECT

KIM'S DETECT PIC, P. 47

Source: Kim (1994)

Neighborhood measures: DETECT

• DETECT similar in spirit to many of the MH, IO/DM measures.

  * localized version of H
  * item-by-item covariation
    — "rest-score splitting"
    —other Rosenbaum-based approaches.

• Can be used to explore for multiple dimensions.

• Nonparametric exploration of multiple dimensions seems to depend on simple structure or approximate simple structure.

  * H-based analysis sorts items into $d = 1$ clusters.
  * Power of DETECT, etc. depends on $d = 1$ clusters.
  * Even in mixed-structure cases, clustering $\Rightarrow d \approx 1$ clusters.

Inference under $d_E = 1$: MLE's

Junker (1991 *Psy*): Look at $d_L = 1$ MLE:

$$\hat{\theta}_J = \arg\max_\theta \prod_{j=1}^{J} P_j(\theta)^{x_j}(1 - P_j(\theta))^{1-x_j}$$

- $\hat{\theta}_J \to \theta$ as $J \to \infty$

- Under appropriate conditions,

$$\sqrt{J}(\hat{\theta}_J - \theta) \sim N\left(0, \frac{\bar{I}_J(\theta) + C_J(\theta)}{\bar{I}_J(\theta)^2}\right),$$

where

$$\bar{I}_J(\theta) = \frac{1}{J}\sum_{j=1}^{J} \lambda_j'(\theta)P_j'(\theta)$$

is the usual test information function,

$$C_J(\theta) = \frac{2}{J}\sum\sum_{1\le i<j\le J} \lambda_i'(\theta)\lambda_j'(\theta)\mathrm{Cov}\,(X_i, X_j|\theta),$$

and $\lambda_j(\theta) = \log P_j(\theta)/[1 - P_j(\theta)]$.

Inference under $d_E = 1$: MLE's

$\widehat{\theta}_J$ asymptotically unbiased, and

$$SE(\widehat{\theta}_J) = \frac{1}{\sqrt{J}}\sqrt{\frac{\overline{I}_J(\theta) + C_J(\theta)}{\overline{I}_J(\theta)^2}};$$

vs. LI-based

$$SE = 1/\sqrt{J \cdot \overline{I}_J(\theta)}$$

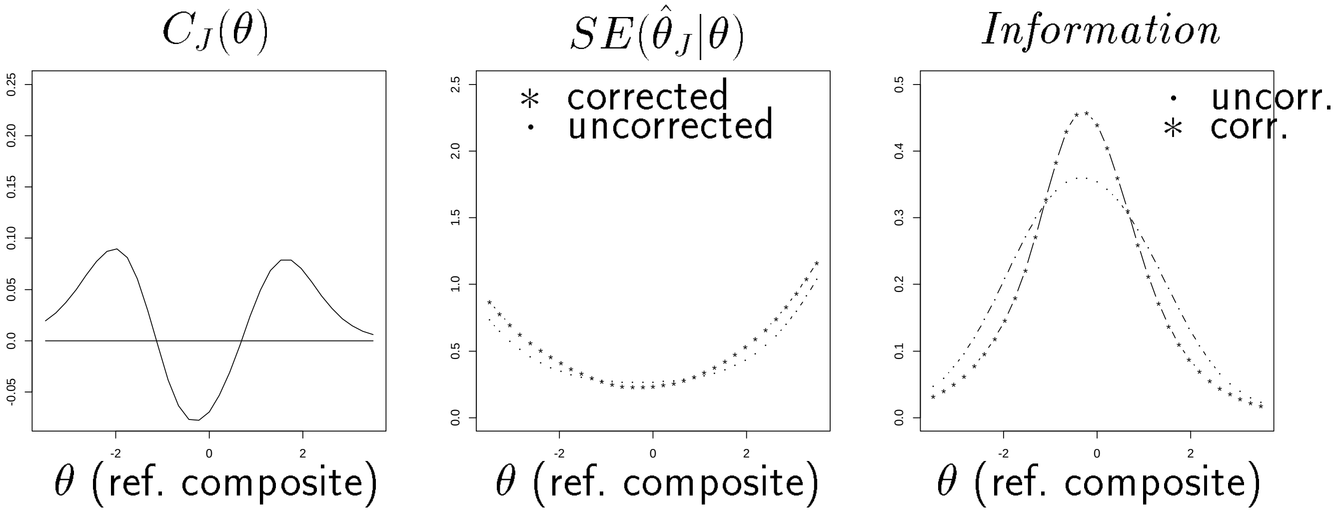
when $C_J(\theta) \equiv 0$.

• In many cases, confidence intervals for θ would be widened.

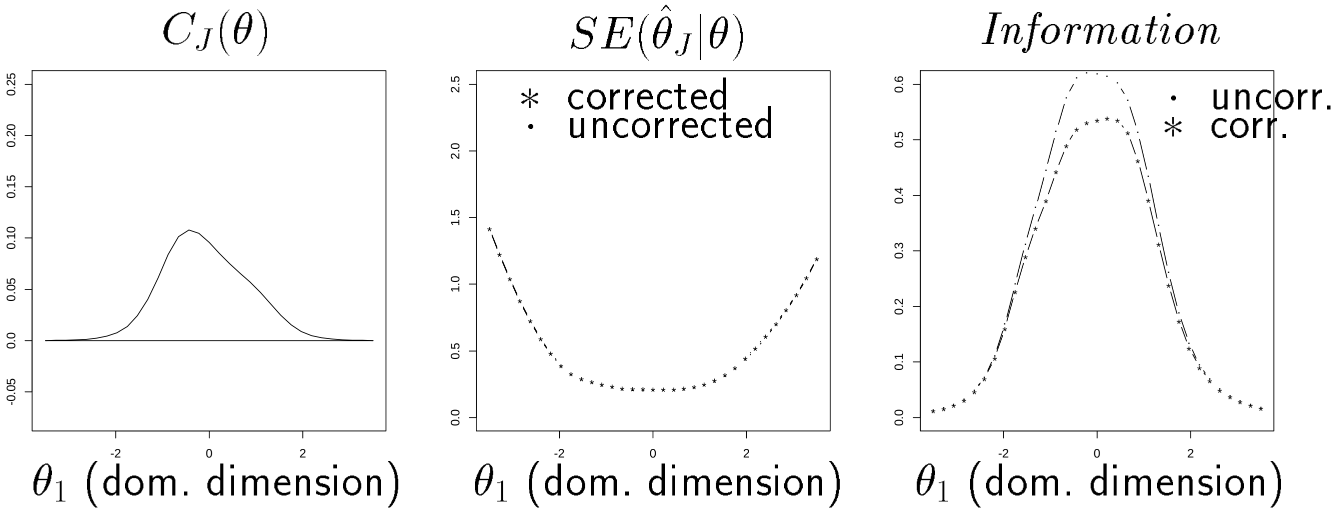
• Possible to *narrow* the CI's by clever test design to make $C_J(\theta) < 0$.

• Might estimate $\widehat{C}_J(\theta)$ "nonparametrically" or from MDIRT estimates (Junker, 1992 *AERA*); (Douglas, Kim and Stout, 1994 *AERA*).

Inference under $d_E = 1$: MLE's



Simple MD2PL: $\rho_{\theta_1,\theta_2} = 0.3$, $\epsilon_J = 2.14$



Paragraph Comprehension: 40 items, 8 paragraphs, $\epsilon_J = 0.05$

Inference under $d_E = 1$: Bayes

Clarke and Junker (1991, *CMU tech rept*):

• The "LI-based posterior"

$$\omega_{LI}(\theta|\underline{x}_J) = \frac{\prod_{j=1}^J P_j(\theta)^{x_j}(1-P_j(\theta))^{1-x_j} f(\theta)}{\int \prod_{j=1}^J P_j(\theta)^{x_j}(1-P_j(\theta))^{1-x_j} f(\theta) d\theta}$$

is asymptotically normal as $J \to \infty$:

$$\mathcal{L}_{\omega_{LI}}\left\{ \frac{\Theta - \widehat{\theta}_J}{\sigma_J} \middle| X_1, \ldots, X_J \right\} \to N(0,1),$$

where

$$\sigma_J = 1/\sqrt{J \cdot \overline{\mathcal{I}}_J(\widehat{\theta}_J)}$$

and

$$\overline{\mathcal{I}}_J(\theta) = -\frac{1}{J}\frac{\partial^2}{\partial\theta^2}\log \prod_{j=1}^J P_j(\theta)^{x_j}(1-P_j(\theta))^{1-x_j}$$

is the usual "observed test information".

• Chang and Stout (1993, *Psy*) have stronger results when LI holds.

Inference under $d_E = 1$: MLE vs. Bayes

Example: "Paragraph Comprehension"

- Binary responses, grouped as X_1, X_2, ..., X_g; X_{g+1}, X_{g+2}, ..., X_{2g}; etc.,

- Identical $P[X_j = 1 \mid \theta] \equiv \theta$.

- $\mathrm{Corr}\,(X_i, X_j|\theta) = \begin{cases} c \text{ if } X_i \text{ and } X_j \text{ in same group,} \\ 0 \text{ if not,} \end{cases}$

Then $d_E = 1$, and

- $\sqrt{J}(\widehat{\theta} - \theta) \approx N(0, \widehat{\theta}(1 - \widehat{\theta})[1 + c(g - 1)])$

- $\sqrt{J}(\Theta - \widehat{\theta})|(X_1, \ldots, X_J) \approx N(0, \widehat{\theta}(1 - \widehat{\theta}))$ □

Inference under $d_E = 1$: MLE vs. Bayes

- Above results generalize to polytomous data

- Both MLE and Bayes consistent, though "LI model is wrong". But...

- Distribution of MLE is asymptotically sensitive to "local dependence" in the data.

  * Rate of concentration of $\hat{\theta}$ around θ computed from the data, not the (wrong) LI model

- Posterior distribution is much less sensitive to "local dependence."

  * Rate of concentration of posterior around $\hat{\theta}$ computed from the (wrong) LI model, not the data.

  * "Proper" Bayesian analysis needs to know the (right) model, to get concentration around θ.

Inference under $d_E = 1$: Nonparametrics

- Kernel estimates of θ distribution:
  Junker (1988 *dissertation*)
  Nandakumar and Junker (1992 *PsySoc*)

- Crude kernel estimates of $P_j(\theta)$ and $C_J(\theta)$
  Junker and Stout (1991, *EMPS*)

- Refined kernel estimates of $P_j(\theta)$ and
  $\text{Cov}\,(X_i, X_j|\theta)$
  Douglas, Kim, Stout (1994 *subm.*)

- Haberman-style joint estimation of $P_j(\theta)$'s
  (by kernel) and θ, as $J \to \infty$
  and $C_1 J^{3/2} \leq N \leq C_2 J^r$.
  LI case: Douglas and Stout (1994 *ms*);

- Motivation: Ramsay (1991 *Psy*; 1993
  TESTGRAF).

Inference under $d_E = 1$: Nonparametrics

Motivation: Est. of $P_j(\theta)$; Ramsay (1991, 93). Careful asymptotics: Douglas and Stout (1994).

1. Obtain crude θ estimates (like $\overline{X}_J$);
2. IRT scale indeterminacy $\Rightarrow$ $\overline{X}_J$ can be transformed to any estimated scale $\tilde{\theta}$ you like.
3. Use a kernel regression approach to estimating ICC's:

$$\widehat{P}_j(\theta) = \frac{\sum_{i=1}^{N} x_{ij} w((\tilde{\theta}_{(i)} - \theta)/h)}{\sum_{n=1}^{N} w((\tilde{\theta}_{(i)} - \theta)/h)}.$$

• Kernel smoothing overcomes discreteness of $\overline{X}_J$ for finite J.

• Must control the rate of smoothing so it is "compatible" with the convergence of $\tilde{\theta}$ [or $\overline{X}_J$] to θ [or $\overline{P}_J(\theta)$].

• Additional theory required since estimated examinee scale scores $\tilde{\theta}_{(i)}$ are <u>random</u> not <u>fixed</u>.

What are we estimating when $d_E > 1$?

Several authors: Unidim. $\widehat{\theta}$ estimates a "linear composite"

$$\theta = \theta_1 \cos\alpha + \theta_2 \sin\alpha$$

when $d = 2$.

• Wang (1986, 87 *ONR*), *Reference composite:*

$\alpha_R \approx$ angle of first princ. component of $(\widehat{\theta}_1, \widehat{\theta}_2)$, depending on MD2PL discriminations.

• Junker and Stout (1991), *Max X_+ information:*

$\alpha_I = \text{argmax}_\alpha\ I(\theta_1 \cos\alpha + \theta_2 \sin\alpha; X_+)$.

• Kim (1994), *Maximum correlation:*

$\alpha_C = \text{argmax}_\alpha\ \rho(\theta_1 \cos\alpha + \theta_2 \sin\alpha, \widehat{\theta})$

* $\alpha_R \approx \alpha_I \approx \alpha_C$ (very worst case: 12° range)

* $\rho(\theta_1 \cos\alpha + \theta_2 \sin\alpha, \widehat{\theta})$ stable as α varies between them.

What are we estimating when $d_E > 1$?

Junker (1992, *AERA*):

- $d_E = 1$: all estimatable composites "converge" to the dominant θ.
- $d_E > 1$: Curvilinear composite, or even lack of consistency is possible.

Example: "Balanced simple structure"

"Full model" IRF's:

$$\pi_j(\theta_1, \theta_2) = P[Z_j < a_{1j}\theta_1 + a_{2j}\theta_2 - b_j]$$

θ_1, θ_2, Z_j's $\sim N(0,1)$
If $\theta = (\theta_1 + \theta_2)/\sqrt{2}$, get "marginal ICC's":

$$P_j(\theta) = P[Z_j < \tilde{a}_j\theta - \tilde{b}_j]$$

where

$$\begin{aligned}
\tilde{a}_j &= (a_{1j} + a_{2j})/\sqrt{2 + (a_{2j} - a_{1j})^2} \\
\tilde{b}_j &= b_j/\sqrt{1 + (a_{2j} - a_{1j})^2/2}.
\end{aligned}$$

Now take $b_j \equiv 0$, and

- Odd items measure θ_1 only;

$$a_{1(2k+1)} = 1 - a_{2(2k+1)} = 1$$

- Even items measure θ_2 only;

$$a_{2(2k)} = 1 - a_{1(2k)} = 1$$

Then

$$P_j(\theta) = P[Z_j < \frac{1}{\sqrt{3}}\theta]$$

and

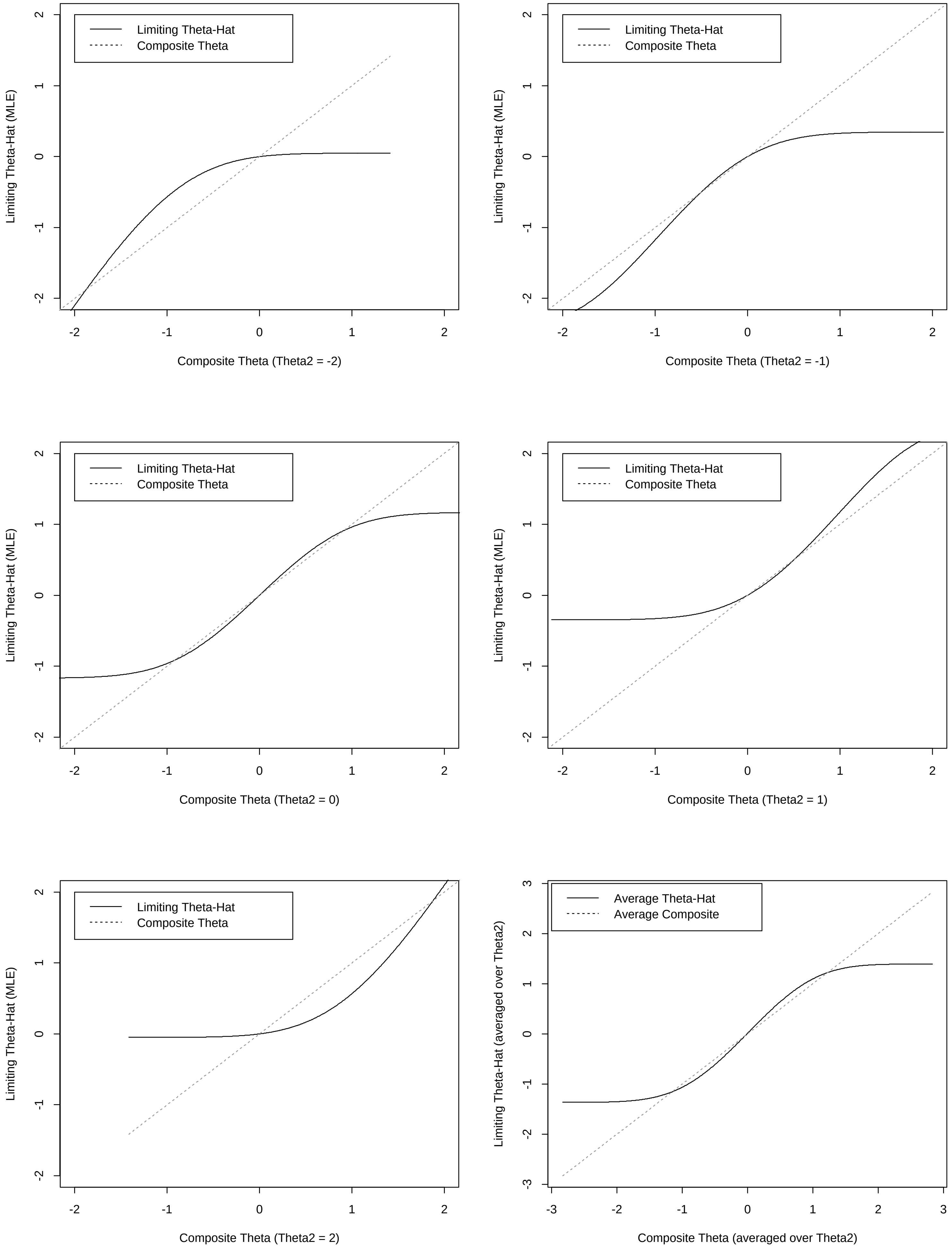
$$\widehat{\theta}_J = \overline{P}_J^{-1}(\overline{X}_J) = \sqrt{3}\Phi^{-1}(\overline{X}_J)$$

where $\Phi(t) = P[Z < t]$.

Since $\overline{X}_J \to (\Phi(\theta_1) + \Phi(\theta_2))/2$, then

$$\widehat{\theta}_J \to \sqrt{3}\Phi^{-1}\left(\frac{\Phi(\theta_1) + \Phi(\theta_2)}{2}\right) \text{ given } (\theta, \theta_2).$$

- $\widehat{\theta}_J \not\to \theta$, or any $f(\theta)$!
- Bias depends on θ_2. $\square$



Limiting Theta-Hat
Composite Theta
Limiting Theta-Hat (MLE)
Composite Theta (Theta2 = -2)
Limiting Theta-Hat
Composite Theta
Limiting Theta-Hat (MLE)
Composite Theta (Theta2 = -1)
Limiting Theta-Hat
Composite Theta
Limiting Theta-Hat (MLE)
Composite Theta (Theta2 = 0)
Limiting Theta-Hat
Composite Theta
Limiting Theta-Hat (MLE)
Composite Theta (Theta2 = 1)
Limiting Theta-Hat
Composite Theta
Limiting Theta-Hat (MLE)
Composite Theta (Theta2 = 2)
Average Theta-Hat
Average Composite
Limiting Theta-Hat (averaged over Theta2)
Composite Theta (averaged over Theta2)

What are we estimating when $d_E > 1$?

Kim (1994) extensive simulation studies:

In "realistic" middle range, estimating a linear average composite ain't bad.

- Compensatory MD2PL and MD3PL models.
- Simple structure, $d = 2$, ranging from 40/0 to 20/20.
- Lowest $\rho(\hat{\theta}, \theta_{\alpha_C}) \approx 0.90$ (MD3PL, 20/20)
- Typical plots:

KIM's PLOT P. 92

Source: Kim (1994)

Conclusions

• $d_E = 1$ is a "robustness testbed" for $d_L = 1$ procedures

• ϵ_J is a good theoretical measure of distance from $d_L = 1$

• Kim's DETECT is a good data-based measure of $d_L = k$

  * Connections with Loevinger's H, etc.
  * Power requires collateral "cluster" info; approximate simple structure—OK!
  * Nonparam. definition of $d = k$, $k > 1$???

• MLE extends well to $d_E = 1$ framework

• Bayes does not extend well to $d_E = 1$

• Nonparametric kernel smoothing extends well; work in progress by Douglas, Stout

• What does unidim. $\widehat{\theta}$ estimate when $d_L > 1$?

  * Theory grim, except when $d_E = 1$;
  * Kim's simulations more hopeful.