Bayesian Checks on Cheating on Tests

Wim J. van der Linden

Charles Lewis

Outline

- Interpretation problems ▷
- Bayesian checks ▷
- Empirical examples ▷
- Discussion ▷

Interpretation Problems

- Current statistical checks based on the idea of hypothesis testing
  - H_0 and H_1 often implicit
  - Null and alternative distributions seldom specified
  - Choice of significance level arbitrary
- These tests have been successful only because of the omnipresence of cheating
  - Any “test” that orders a population w.r.t. a relevant statistic would have been!

Interpretation Problems

- Schools don’t understand hypothesis testing
  - Tradeoff between Type 1 and II errors
  - Significant ≠ true
- How to account for a known proportion of the population that actually cheats?
- Statistical tests at different levels
  - Students, classes, schools
  - Results may seem inconsistent (statistically they are not!)

Interpretation Problems

- Should we condition? If so, how?
  - *K*-index: *incorrect* responses by *s*
  - Generalized binomial test: *none* of the responses by *s*
  - Conditional version of same test: *all* of responses by *s*
- Expected power of conditional test is equal to power of unconditional test (Lehmann & Romano, 2005, chap. 10)

Interpretation Problems *Cont'd*

- Only nonstatistical reasons to choose an unconditional test over a conditional one, e.g.,
  - repetitive use of test
  - symmetry
  - computational reasons

Bayesian Checks

- Bayesian approach
  - Posterior odds of cheating easier to interpret than significance probabilities
  - Automatic allows for known proportion that actually cheats (prior probabilities)
  - Conditioning on *all* responses, *both by the source and the copier*

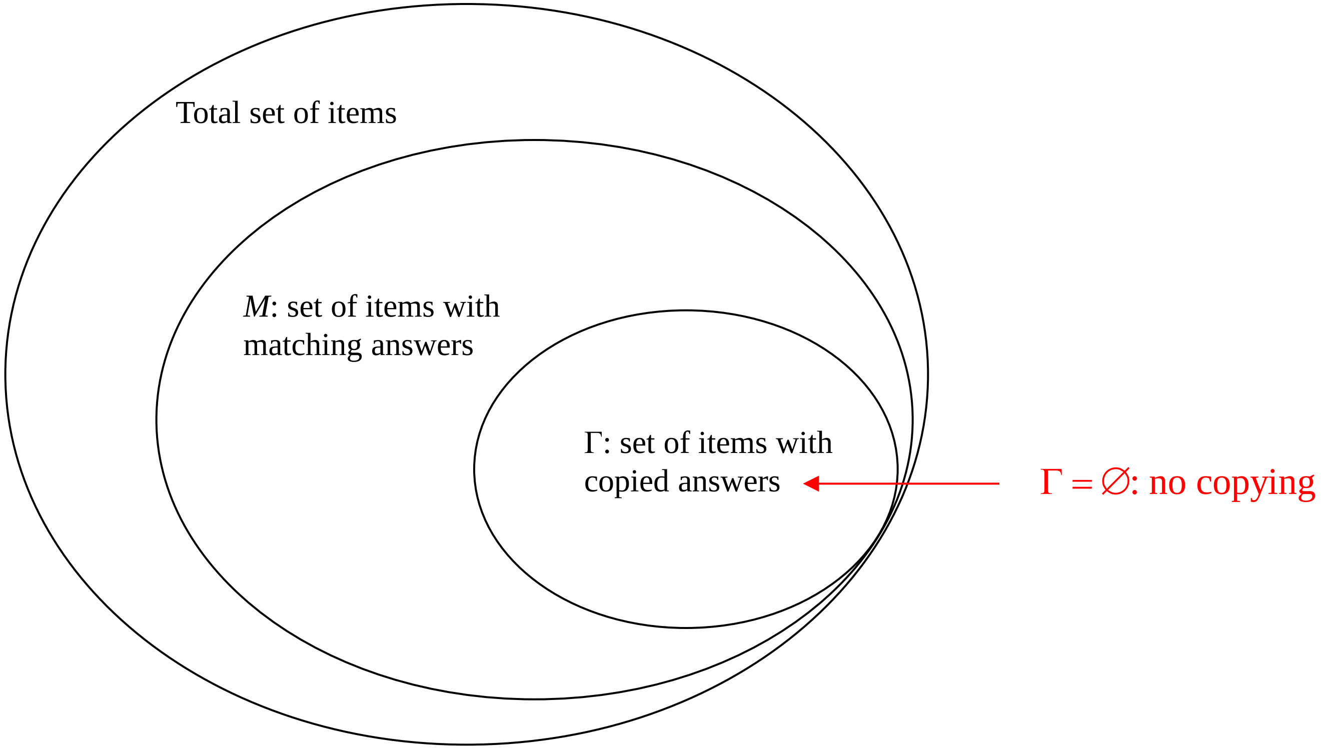
Bayesian Checks *Cont'd*

- Checks should only be used if there is prior evidence suggesting scrutiny of a specific portion of the test, e.g., a section in the test or page of the answer sheet
  - Blind applications to the complete tests for all pairs of test takers are meaningless
- Example for check on answer copying
  - For other types of cheating, see full paper

Bayesian Checks *Cont'd*

- Notation
  - c: hypothetical copier
  - s: source
  - M: set of all items with matching (correct or incorrect) responses
  - $\Gamma \subseteq M$: (unknown) subset of M actually copied
  - $\Gamma = \varnothing$: no copying
  - $p(\Gamma)$: prior probability of subset Γ being copied (defined over all possible subsets of M)

Bayesian Checks *Cont'd*



Bayesian Checks *Cont'd*

- Response probability for copier on item i

$$p(y_{ci}|\theta_c,\Gamma,y_{si}) = \begin{cases} p(y_{ci}|\theta_c), & \text{if } i \notin \Gamma \\ 1, & \text{if } i \in \Gamma, \quad y_{ci} = y_{si} \\ 0, & \text{if } i \in \Gamma, \quad y_{ci} \neq y_{si} \end{cases}$$

- Probability of response vectors $\mathbf{y}_c$ and $\mathbf{y}_s$

$$p(\mathbf{y}_c,\mathbf{y}_s|\theta_c,\theta_s,\Gamma) = p(\mathbf{y}_c|\theta_c,\Gamma,\mathbf{y}_s)\,p(\mathbf{y}_s|\theta_s)$$

$$= \prod_{i=1}^{n} p(y_{ci}|\theta_c,\Gamma,y_{si}) \prod_{i=1}^{n} p(y_{si}|\theta_s)$$

Bayesian Checks *Cont'd*

- Response probability for copier on item *i*

$$p(y_{ci} | \theta_c, \Gamma, y_{si}) = \begin{cases} p(y_{ci} | \theta_c), & \text{if } i \notin \Gamma \\ 1, & \text{if } i \in \Gamma, \quad y_{ci} = y_{si} \\ 0, & \text{if } i \in \Gamma, \quad y_{ci} \neq y_{si} \end{cases}$$

- Probability of response vectors $\mathbf{y}_c$ and $\mathbf{y}_s$

$$p(\mathbf{y}_c, \mathbf{y}_s | \theta_c, \theta_s, \Gamma) = p(\mathbf{y}_c | \theta_c, \Gamma, \mathbf{y}_s) p(\mathbf{y}_s | \theta_s)$$

$$= \prod_{i=1}^{n} p(y_{ci} | \theta_c, \Gamma, y_{si}) \prod_{i=1}^{n} p(y_{si} | \theta_s)$$

Bayesian Checks *Cont'd*

- Posterior probability of $\Gamma = \varnothing$ (i.e., no copying)

$$p(\varnothing | \mathbf{y}_c, \mathbf{y}_s) = \frac{p(\varnothing)\prod_{i=1}^{n} p(y_{ci} | \theta_c, \Gamma, y_{si})\prod_{i=1}^{n} p(y_{si} | \theta_s)}{\sum_{\Gamma} p(\Gamma)\prod_{i=1}^{n} p(y_{ci} | \theta_c, \Gamma, y_{si})\prod_{i=1}^{n} p(y_{si} | \theta_s)}$$

$$= \frac{p(\varnothing)\prod_{i=1}^{n} p(y_{ci} | \theta_c, \Gamma, y_{si})}{\sum_{\Gamma} p(\Gamma)\prod_{i=1}^{n} p(y_{ci} | \theta_c, \Gamma, y_{si})}$$

Bayesian Checks *Cont'd*

- Posterior probability of $\Gamma = \varnothing$ (*cont'd*)

$$p(\varnothing \mid \mathbf{y}_c, \mathbf{y}_s) = \frac{p(\varnothing)\prod_{i\in M} p(y_{ci} \mid \theta_c)\prod_{i\in \overline{M}} p(y_{ci} \mid \theta_c)}{\sum_{\Gamma\subseteq M} p(\Gamma)\prod_{i\in M} p(y_{ci} \mid \theta_c, \Gamma, y_{si})\prod_{i\in \overline{M}} p(y_{ci} \mid \theta_c)}$$

$$= \frac{p(\varnothing)\prod_{i\in M} p(y_{ci} \mid \theta_c)}{\sum_{\Gamma\subseteq M} p(\Gamma)\prod_{i\in M} p(y_{ci} \mid \theta_c, \Gamma, y_{si})}$$

Bayesian Checks *Cont'd*

- Posterior odds of cheating can be shown to simplify to

$$\frac{1-p(\varnothing|\mathbf{y}_c,\mathbf{y}_s)}{p(\varnothing|\mathbf{y}_c,\mathbf{y}_s)} = \frac{\sum_{\substack{\Gamma\neq\varnothing \\ \Gamma\subseteq M}} p(\Gamma)\prod_{i\in M\setminus\Gamma} p(y_{ci}|\theta_c)}{p(\varnothing)\prod_{i\in M} p(y_{ci}|\theta_c)}$$

- Thus, odds are independent of
  - ability of source
  - any responses outside set of matches, *M*

Bayesian Checks *Cont'd*

- Proposed specification of prior probabilities of cheating
  - Choice of prior probability of no cheating, $p(\varnothing)$, (or odds of no cheating)
  - Assumption of independence across items
  - Choice of prior probabilities γ_i of cheating on item i to be consistent with

$$p(\varnothing) = \prod_{i \in N} (1 - \gamma_i)$$

Bayesian Checks *Cont'd*

- Posterior odds of cheating are now equal to

$$\frac{1-p(\varnothing \mid \mathbf{y}_c, \mathbf{y}_s)}{p(\varnothing \mid \mathbf{y}_c, \mathbf{y}_s)} = \frac{\sum_{\substack{\Gamma \neq \varnothing \\ \Gamma \subseteq M}} \prod_{i \in \Gamma} \gamma_i \prod_{i \in M \setminus \Gamma} (1-\gamma_i) p_{ci}}{\prod_{i \in M} (1-\gamma_i) p_{ci}}$$

- Complicated combinatorial expression
  - Example for set of 3 items
  - Notation: $\xi_{ci}=(1-\gamma_i)p_{ci}$ (compare with $\gamma_i \cdot 1$)

Posterior Odds for Set of 3 Items

Items	Products of γ_i and ξ_{ci}		
	γ	ξ	$\gamma'\xi$
∅	1	$\xi_{c1}\xi_{c2}\xi_{c3}$	$\xi_{c1}\xi_{c2}\xi_{c3}$
{1}	γ_1	$\xi_{c2}\xi_{c3}$	$\gamma_1\xi_{c2}\xi_{c3}$
{2}	γ_2	$\xi_{c1}\xi_{c3}$	$\xi_{c1}\gamma_2\xi_{c3}$
{3}	γ_3	$\xi_{c1}\xi_{c2}$	$\xi_{c1}\xi_{c2}\gamma_{c3}$
{1,2}	$\gamma_1\gamma_2$	ξ_{c3}	$\gamma_1\gamma_2\xi_{c3}$
{1,3}	$\gamma_1\gamma_3$	ξ_{c2}	$\gamma_1\xi_{c2}\gamma_3$
{2,3}	$\gamma_2\gamma_3$	ξ_{c1}	$\xi_{c1}\gamma_2\gamma_3$
{1,2,3}	$\gamma_1\gamma_2\gamma_3$	1	$\gamma_1\gamma_2\gamma_3$

Denominator (row ∅)

Numerator (rows {1} through {1,2,3})

Bayesian Checks *Cont'd*

- Odds can be calculated using modified version of well-known algorithm for calculation of number-correct score distributions ("Lord-Wingersky algorithm")
  - Treat ξ_{ci} as probability of correct response on i
  - Treat γ_i as probability of incorrect response on i

    (They are no probabilities, though!)

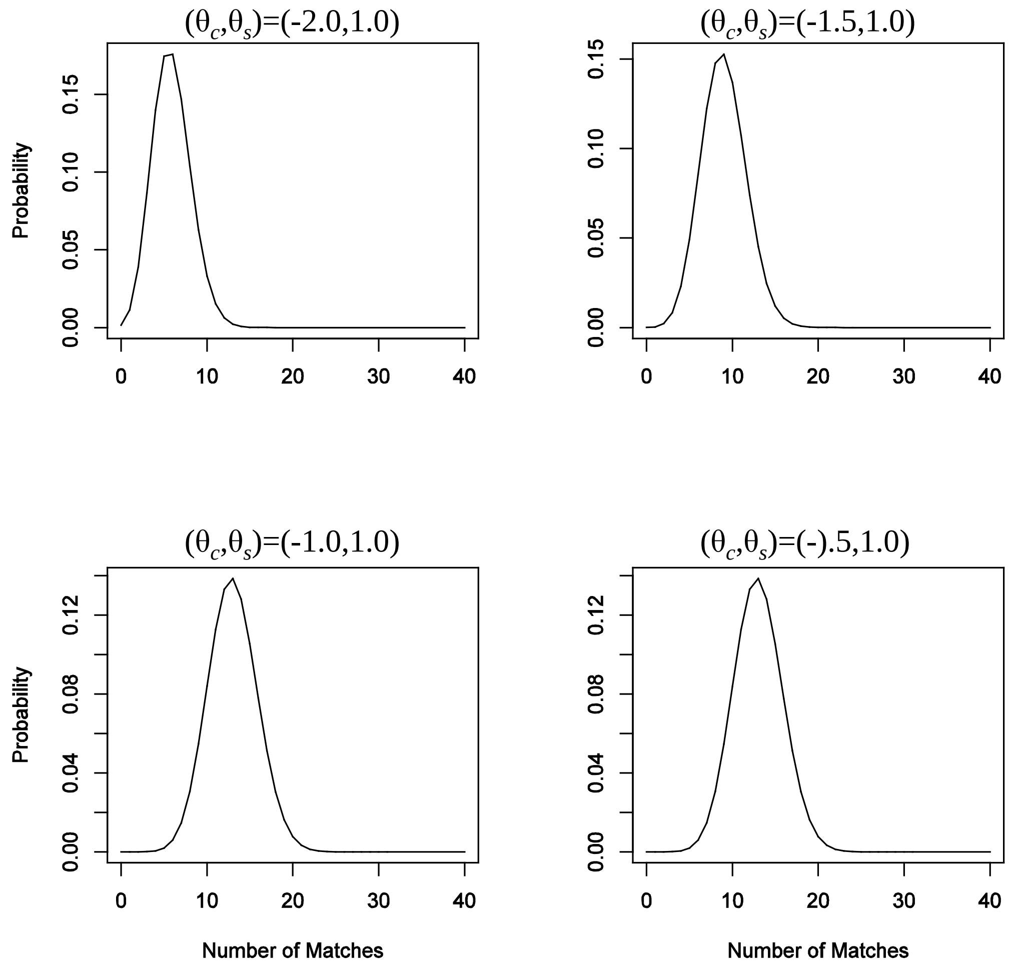
Empirical Examples

- NRM parameters for 40-item test with 5 answer choices on each item (Wollack, 1997)
- Suppose a proctor has witnessed supicious communication between *c* and *s* during first section of 10 items)
- Prior probability of cheating on at least one item: $1-p(\varnothing)=.25,\ .50,\ \text{and}\ .75$
  - That is, odds of cheating equal to 1:3, 1:1, and 3:1

Empirical Examples

- No prior knowledge as to which items in section were involved: $\gamma = 1 - [1 - p(\varnothing)]^{-10}$
- Four pairs of abilities:
  - θ_c=-2.0, -1.5, -1.0, and -0.5
  - θ_s=1.0
- Answer copying simulated by replacing copier's responses by those by source in the data file

Distribution of Number of Random Matches on Entire Test



Posterior Odds for Different Cases

(θ_c,θ_s)	(-2.0,1.0)			(-1.5,1.0)		
Prior Odds	1:3	1:1	3:1	1:3	1:1	3:1
No. of Matches						
2	0.62	1.71	4.28	0.31	0.82	1.98
3	2.04	7.56	27.58	0.41	1.18	3.18
4	2.42	10.51	45.50	0.76	2.52	8.50
5	3.15	16.00	444.05	0.90	3.20	12.31
6	6.06	45.44	681.63	1.62	7.11	38.02
7	6.82	57.40	*	4.48	28.91	255.20
8	13.96	188.98	*	5.12	37.47	407.13
9	15.59	239.49	*	12.89	157.66	*
10	19.85	391.38	*	15.11	219.86	*

Posterior Odds for Different Cases

(θ_c,θ_s)	**(-1.0,1.0)**			**(-0.5,1.0)**		
Prior Odds	1:3	1:1	3:1	1:3	1:1	3:1
No. of Matches						
2	0.19	0.50	1.15	0.12	0.31	0.69
3	0.67	1.99	5.56	0.20	0.53	1.27
4	0.98	3.36	11.78	0.34	0.98	2.65
5	1.14	4.18	16.82	0.47	1.43	4.38
6	1.38	5.65	27.56	0.63	2011	7.50
7	1.63	7.36	42.35	0.75	2.65	10.55
8	8.97	64.74	658.47	0.85	3.18	14.02
9	9.56	74.34	858.08	0.94	3.68	17.70
10	11.10	101.38	*	1.12	4.72	26.35

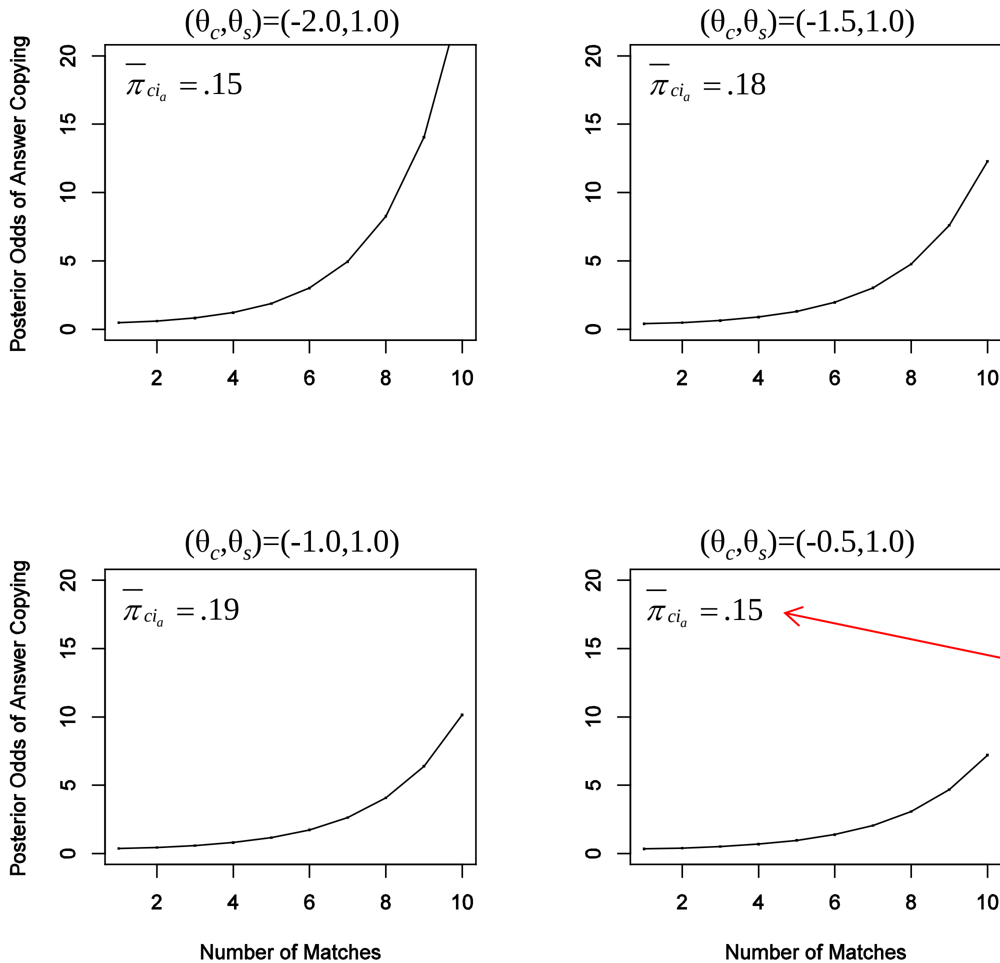
Discussion

- Posterior odds
  - – increase with number of matches
  - – increase with prior odds,
  - – but decrease with difference between θ_c and θ_s
- Posterior odds are strongly data dependent
  - – Same number of matches but on different items and/or alternatives leads to different odds

Discussion *Cont'd*

- Two examples in which we ignore information in the data set
  - differences between items
  - alternatives chosen by source
- Again, Bayesian checks are only for specific hypothesis on specific parts of the test
  - More informative prior distributions
  - Possible to estimate θ_c from remaining portion of the test

Posterior Odds as Function of Number of Matches



Average conditional probability of random match given choice by source

Posterior Odds as Function of Number of Matches