EDUCATIONAL and PSYCHOLOGICAL MEASUREMENT

VOLUME 82 NUMBER 5 OCTOBER 2022

Application of Change Point Analysis of Response Time Data to Detect Test Speededness

Ying Cheng

University of Notre Dame, Notre Dame, IN, USA

Can Shao

Applied Materials Inc, Santa Clara, CA, USA

Reporter: Yingshi Huang

Introduction

- Response time data
  - **collection:** computer-based testing and online survey
  - **advantages:** improve item parameter estimation & test assembly

- Detection of aberrant response behavior
  - **speededness**
  - low motivation or lack of effort
  - item pre-knowledge

Introduction

- What is speededness?

- How does it influence the test results?

- How did previous researchers remedy this problem?

Introduction

- What is speededness?
  - time limit affects examinees' performance (speed is not part of the test purports)
  - unreached items or rapid guessing behavior
- How does it influence the test results?
- How did previous researchers remedy this problem?

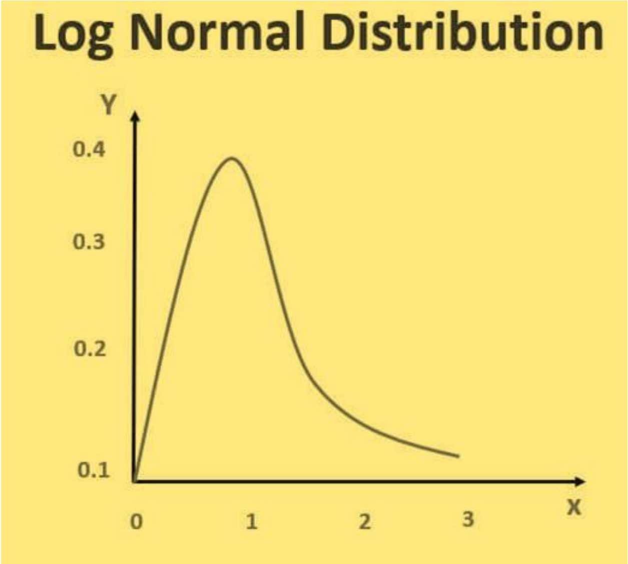
Introduction

- What is speededness?
  - time limit affects examinees' performance (speed is not part of the test purports)
  - unreached items or rapid guessing behavior
- How does it influence the test results?
  - lower response accuracy or missing responses
  - bias the estimates of item and ability estimates
  - distort the relationship between test scores and other variables
- How did previous researchers remedy this problem?

Introduction

- What is speededness?
  - time limit affects examinees' performance (speed is not part of the test purports)
  - unreached items or rapid guessing behavior
- How does it influence the test results?
  - lower response accuracy or missing responses
  - bias the estimates of item and ability estimates
  - distort the relationship between test scores and other vari
- How did previous researchers remedy this problem?
  - model the distribution of response time

    fit? or **not fit?** (with large residuals)



Introduction

- What is speededness?
  - time limit affects examinees’ performance (speed is not part of the test purports)
  - unreached items or rapid guessing behavior
- How does it influence the test results?
  - lower response accuracy or missing responses
  - bias the estimates of item and ability estimates
  - distort the relationship between test scores and other variables
- How did previous researchers remedy this problem?
  - model the distribution of response time
  - model response time data that are affected by test speededness

  mixture modeling $\log(T_{ij}) \sim N(\mu_c, \sigma_c^2)$ $\log(T_{ij}) \sim N(\beta_j - \tau_i, \sigma_j^2)$

Introduction

- What is speededness?
  - time limit affects examinees' performance (speed is not part of the test purports)
  - unreached items or rapid guessing behavior
- How does it influence the test results?
  - lower response accuracy or missing responses
  - bias the estimates of item and ability estimates
  - distort the relationship between test scores and other variables
- How did previous researchers remedy this problem?
  - model the distribution of response time
  - model response time data that are affected by test speededness
  - visual inspection or arbitrarily chosen thresholds: **e.g., 3 second threshold**

Introduction

- What is speededness?
  - time limit affects examinees' performance (speed is not part of the test purports)
  - unreached items or rapid guessing behavior
- How does it influence the test results?
  - lower response accuracy or missing responses
  - bias the estimates of item and ability estimates
  - distort the relationship between test scores and other
- How did previous researchers remedy this problem?
  - model the distribution of response time
  - model response time data that are affected by test speededness
  - visual inspection or arbitrarily chosen thresholds

What is still unknown?

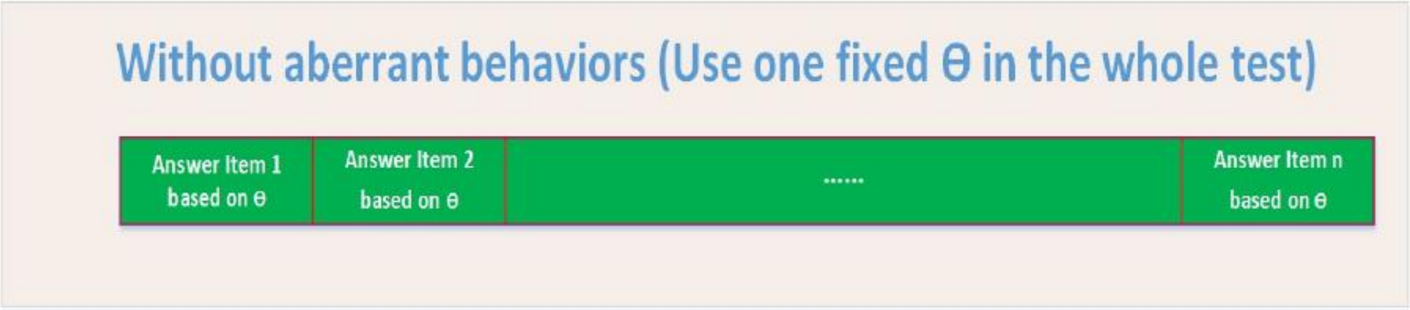
<u>the speeding point</u>

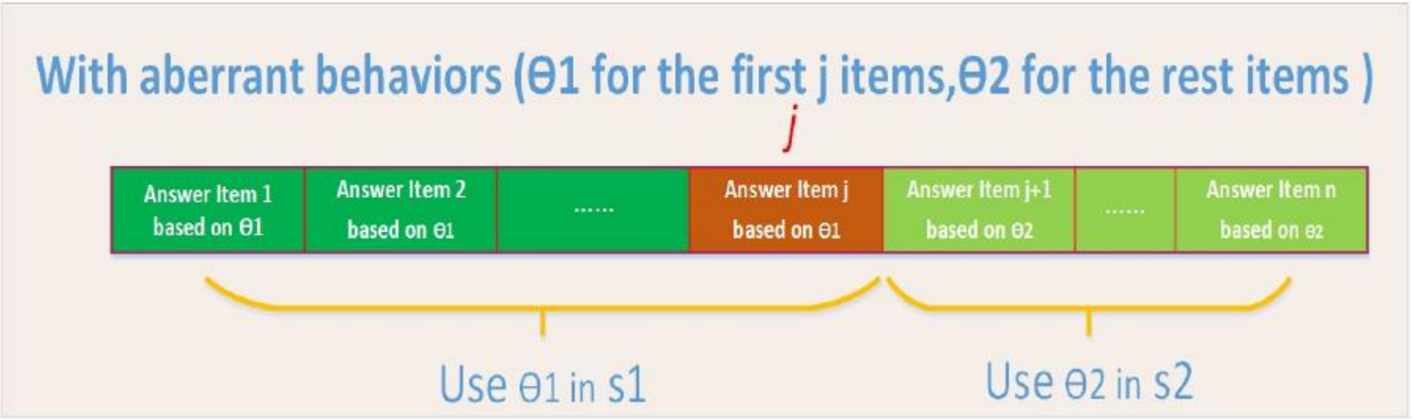
Purpose:

propose a procedure based on change point analysis (CPA)

Method

- Change point analysis (CPA)
  - test taker *i* operates under time pressure after a specific item





From Xiaofeng Yu's presentation

Method

- CPA for the detection of speededness

PSYCHOMETRIKA—VOL. 81, NO. 4, 1118–1141
DECEMBER 2016
DOI: 10.1007/s11336-015-9476-7

Psychometric Society

CrossMark

DETECTION OF TEST SPEEDEDNESS USING CHANGE-POINT ANALYSIS

CAN SHAO, JUN LI, AND YING CHENG

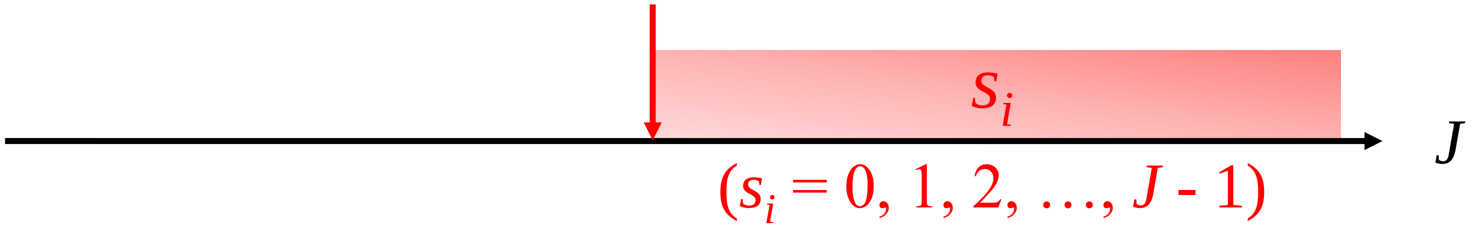
UNIVERSITY OF NOTRE DAME

Change-point analysis (CPA) is a well-established statistical method to detect abrupt changes, if any, in a sequence of data. In this paper, we propose a procedure based on CPA to detect test speededness. This procedure is not only able to classify examinees into speeded and non-speeded groups, but also identify the point at which an examinee starts to speed. Identification of the change point can be very useful. First, it informs decision makers of the appropriate length of a test. Second, by removing the speeded responses, instead of the entire response sequence of an examinee suspected of speededness, ability estimation can be improved. Simulation studies show that this procedure is efficient in detecting both speeded examinees and the speeding point. Ability estimation is dramatically improved by removing speeded responses identified by our procedure. The procedure is then applied to a real dataset for illustration purpose.

Key words: test speededness, change-point analysis, false discovery rate, likelihood ratio statistic, item response theory.

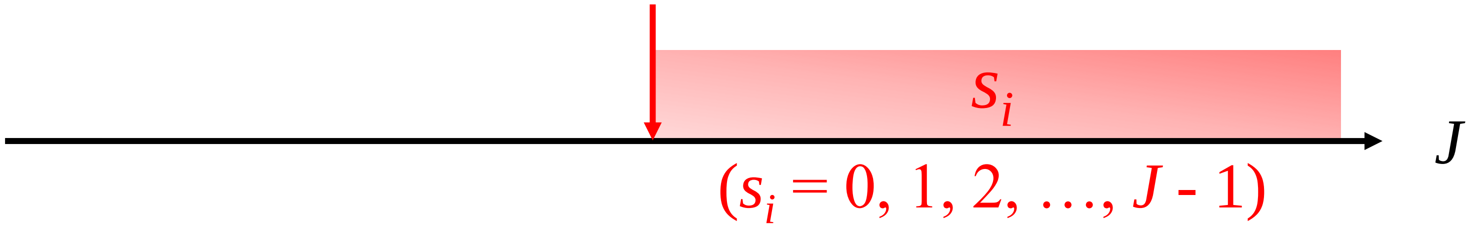
Method

- CPA for the detection of speededness: consider examinee i



Method

- CPA for the detection of speededness: consider examinee *i*



  - it is natural to assume the examinee's ability to drop (d_i) as he or she starts to speed

$$P_{ij}(\theta) = \frac{\exp\left[a_j\left(\theta_i - b_j - \boxed{d_i} \cdot I\left(j > J - s_i\right)\right)\right]}{1 + \exp\left[a_j\left(\theta_i - b_j - d_i \cdot I\left(j > J - s_i\right)\right)\right]}$$

Shao, Li, & Cheng, 2016 *PSYCHOMETRIKA*

Method

- CPA for the detection of speededness: consider examinee i
  - **the goal is to** test the existence of the speed point and pinpoint its location

$$H_0 : s_i = 0,$$
$$H_a : s_i > 0.$$

the speed point do not exist

the speed point exists

Method

- CPA for the detection of speededness: consider examinee i
  - **the goal is to** test the existence of the speed point and pinpoint its location

$$H_0 : s_i = 0, \qquad \text{the speed point do not exist}$$
$$H_a : s_i > 0. \qquad \text{the speed point exists}$$

1. compute the distance of the log-likelihood under H_0 and H_a:

$$\Delta l_i = 2\left(l_i^{Ha} - l_i^{H0}\right)$$

2. construct the null distribution of Δl_i
3. find a cutoff for the *p* value (this article used FDR to correct multiple comparisons)
4. compare the observed Δl_i against the cutoff value

Shao, Li, & Cheng, 2016 *PSYCHOMETRIKA*

Method

- CPA for the detection of speededness: consider examinee i
  - **the goal is to** test the existence of the speed point and pinpoint its location

$$H_0 : s_i = 0,$$
$$H_a : s_i > 0.$$

→ the speed point do not exist

the speed point exists

1. compute the distance of the log-likelihood under H_0 and H_a:

$$\Delta l_i = 2\left(l_i^{Ha} - l_i^{H0}\right)$$

2. construct the null distribution of Δl_i
3. find a cutoff for the *p* value (this article used FDR to correct multiple comparisons)
4. compare the observed Δl_i against the cutoff value

Shao, Li, & Cheng, 2016 *PSYCHOMETRIKA*

Method

- CPA for the detection of speededness: consider examinee i

1. compute the distance of the log-likelihood under H_0 and H_a:

$$\Delta l_i = 2\left(l_i^{Ha} - l_i^{H0}\right)$$

Method

- CPA for the detection of speededness: consider examinee i
  1. compute the distance of the log-likelihood under H_0 and H_a:

$$\Delta l_i = 2\left(l_i^{Ha} - l_i^{H0}\right)$$

suppose $s_i = k > 0$

$$\checkmark\ P_{ij}(\theta) = \begin{cases} \frac{\exp[a_j(\theta_i - b_j)]}{1+\exp[a_j(\theta_i - b_j)]}, & if\ j \leq (J-k) \\ \frac{\exp[a_j(\theta_i - b_j - d_i)]}{1+\exp[a_j(\theta_i - b_j - d_i)]}, & \text{otherwise} \end{cases}$$

normal responses: θ_i

speeded responses: $\theta_i - d_i$

(be bounded between −4.0 and +4.0)

Method

- CPA for the detection of speededness: consider examinee i
  1. compute the distance of the log-likelihood under H_0 and H_a:

$$\Delta l_i = 2\left(l_i^{Ha} - l_i^{H0}\right)$$

suppose $s_i = k > 0$

✓ $$P_{ij}(\theta) = \begin{cases} \frac{\exp[a_j(\theta_i - b_j)]}{1+\exp[a_j(\theta_i - b_j)]}, & if\ j \le (J-k) \\ \frac{\exp[a_j(\theta_i - b_j - d_i)]}{1+\exp[a_j(\theta_i - b_j - d_i)]}, & \text{otherwise} \end{cases}$$

→ normal responses: θ_i

speeded responses: $\theta_i - d_i$

(be bounded between −4.0 and +4.0)

✓ $$l_i^{(k)} = \sum_{j=1}^{J}\left[u_{ij}\ln P_{ij}^{(k)} + (1-u_{ij})\ln Q_{ij}^{(k)}\right]$$

Shao, Li, & Cheng, 2016 *PSYCHOMETRIKA*

Method

- CPA for the detection of speededness: consider examinee i
  1. compute the distance of the log-likelihood under H_0 and H_a:

$$\Delta l_i = 2\left(l_i^{Ha} - l_i^{H0}\right)$$

suppose $s_i = k > 0$

$$\checkmark\ P_{ij}(\theta) = \begin{cases} \frac{\exp[a_j(\theta_i - b_j)]}{1+\exp[a_j(\theta_i - b_j)]}, & if\ j \le (J-k) \\ \frac{\exp[a_j(\theta_i - b_j - d_i)]}{1+\exp[a_j(\theta_i - b_j - d_i)]}, & \text{otherwise} \end{cases}$$

normal responses: θ_i

speeded responses: $\theta_i - d_i$

(be bounded between −4.0 and +4.0)

$$\checkmark\ l_i^{(k)} = \sum_{j=1}^{J}\left[u_{ij}\ln P_{ij}^{(k)} + (1-u_{ij})\ln Q_{ij}^{(k)}\right]$$

Shao, Li, & Cheng, 2016 *PSYCHOMETRIKA*

Method

- CPA for the detection of speededness: consider examinee i
  1. compute the distance of the log-likelihood under H_0 and H_a:

✓ $$l_i^{(k)} = \sum_{j=1}^{J} \left[u_{ij} \ln P_{ij}^{(k)} + \left(1 - u_{ij}\right) \ln Q_{ij}^{(k)} \right]$$

➡ find the desired k value that can maximize the log-likelihood

$$\hat{s}_i = \arg\max_{k=1,2,\ldots,(J-1)} \left\{ l_i^{(k)} \right\}$$

$$l_i^{Ha} = \max_{k=1,2,\ldots,(J-1)} \left\{ l_i^{(k)} \right\}$$

Shao, Li, & Cheng, 2016 *PSYCHOMETRIKA*

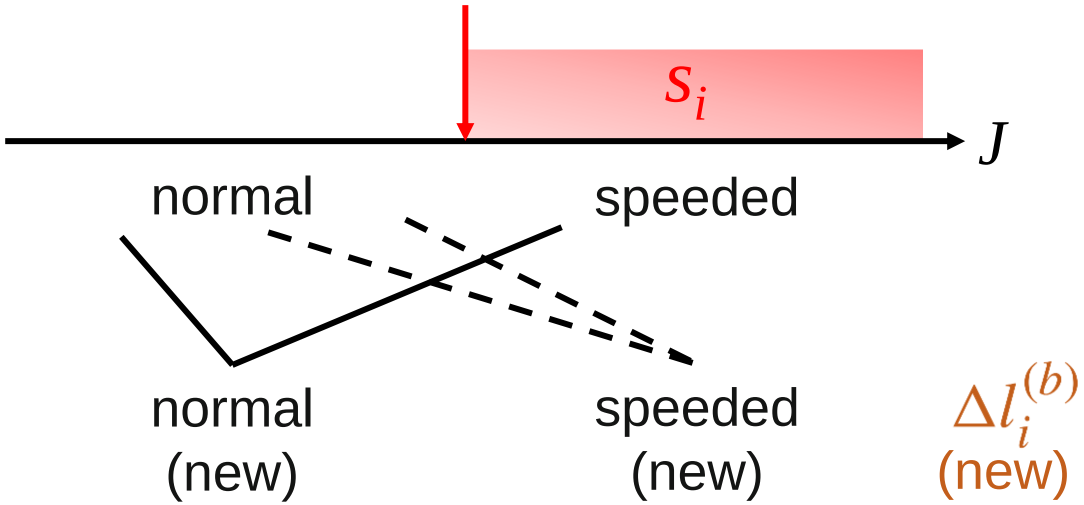
Method

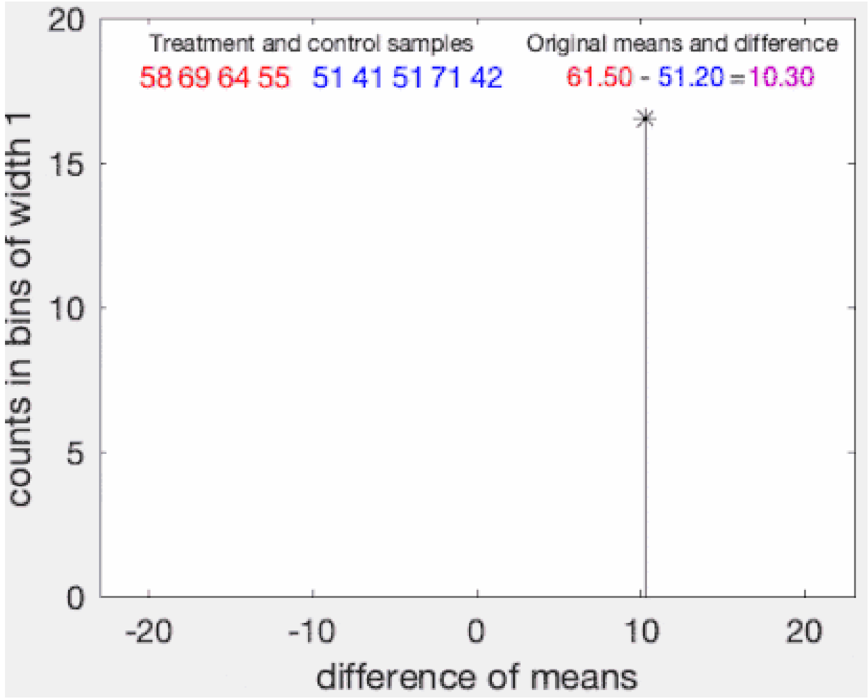
- CPA for the detection of speededness: consider examinee i
  2. construct the null distribution of Δl_i (with no change point)
     - ✓ **challenge:** does not follow a χ^2 distribution with a known degrees of freedom
     - ✓ **solution:** permutation test

Method

- CPA for the detection of speededness: consider examinee i
  2. construct the null distribution of Δl_i (with no change point)
     - ✓ **challenge:** does not follow a χ^2 distribution with a known degrees of freedom
     - ✓ **solution:** permutation test

randomly permute the item responses





From Wikipedia

Shao, Li, & Cheng, 2016 *PSYCHOMETRIKA*

Method

- Shao et al. (2016)
  - assuming two separate abilities
  - computing the maximum change in likelihood
  - performing the significant testing

Method

- Shao et al. (2016)
  - assuming two separate abilities
  - computing the maximum change in likelihood
  - performing the significant testing

- Rooms for improvement
  - relying solely on the dichotomous item response data
  - reliance on the permutation test: computationally cumbersome (sample size & test length)

Method

- Shao et al. (2016)
  - assuming two separate abilities
  - computing the maximum change in likelihood
  - performing the significant testing

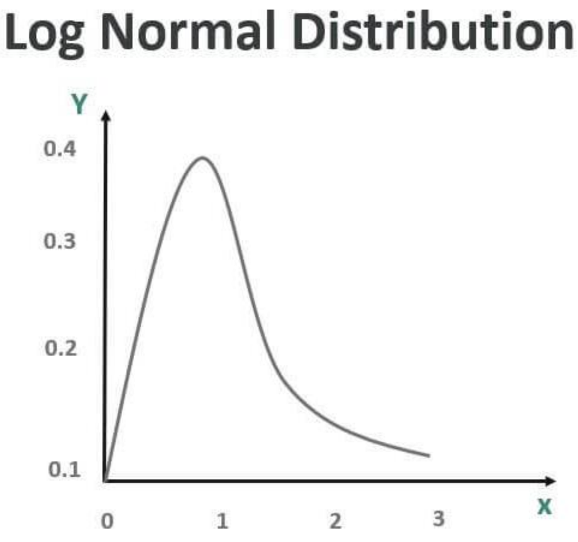
- Rooms for improvement
  - relying solely on the dichotomous item response data
  - reliance on the permutation test: computationally cumbersome (sample size & test length)

→

1. using continuous response time data
2. establishing generally applicable cutoffs
3. identifying factors that influence the estimation of change point

Method

- CPA for item response time data

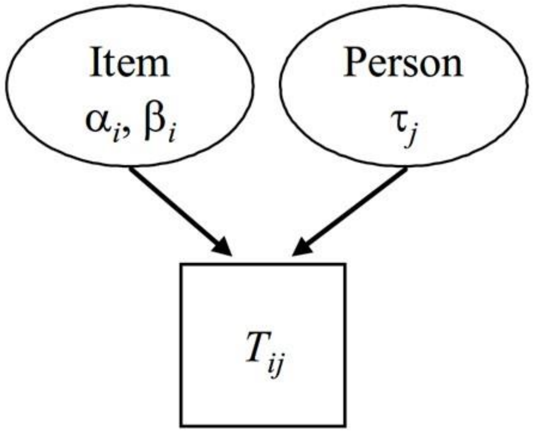


discrimination a ability θ

$$f(t_{ij};\tau_i,\alpha_j,\beta_j)=\frac{\alpha_j}{t_{ij}\sqrt{2\pi}}exp\{-\frac{1}{2}[\alpha_j(\ln t_{ij}-(\beta_j-\tau_i))]^2\}$$

difficulty b

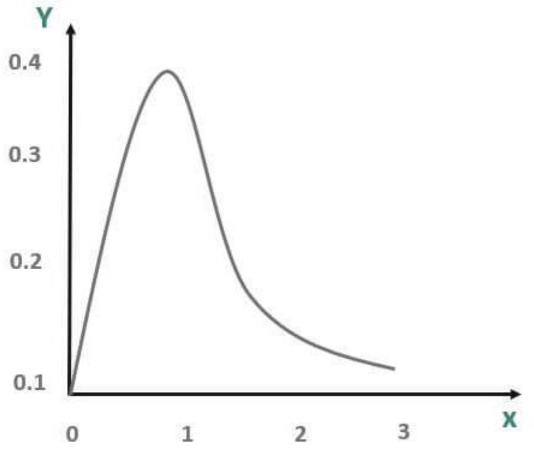
$$\ln(t_{ij})=\beta_j-\tau_i+\varepsilon_{ij},\ \ \varepsilon_{ij}\sim N(0,\alpha_j^{-2})$$



Method

- CPA for item response time data

Log Normal Distribution

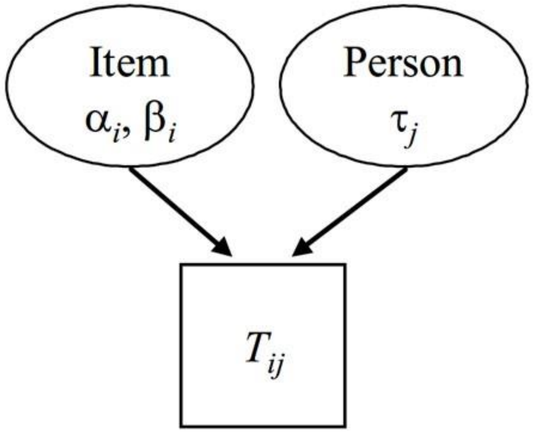


discrimination a ability θ

$$f(t_{ij};\tau_i,\alpha_j,\beta_j)=\frac{\alpha_j}{t_{ij}\sqrt{2\pi}}exp\{-\frac{1}{2}[\alpha_j(\ln t_{ij}-(\beta_j-\tau_i))]^2\}$$

difficulty b

$$\ln(t_{ij})=\beta_j-\tau_i+\varepsilon_{ij},\ \ \varepsilon_{ij}\sim N(0,\alpha_j^{-2})$$



Two CPA test statistics:

1. the Likelihood Ratio Test
2. the Wald Test

Method

- CPA for item response time data: the likelihood ratio test

$$f(t_{ij};\tau_i,\alpha_j,\beta_j)=\frac{\alpha_j}{t_{ij}\sqrt{2\pi}}exp\{-\frac{1}{2}[\alpha_j(\ln t_{ij}-(\beta_j-\tau_i))]^2\}$$

$$L(\tau_i;\mathbf{t_i})=\prod_{j=1}^{J}\frac{\alpha_j}{t_{ij}\sqrt{2\pi}}exp\{-\frac{1}{2}[\alpha_j(\ln t_{ij}-(\beta_j-\tau_i))]^2\}$$

$$l(\tau_i;\mathbf{t_i})=lnL(\tau_i;\mathbf{t_i})=\sum_{j=1}^{J}\ln\frac{\alpha_j}{t_{ij}\sqrt{2\pi}}-\frac{1}{2}\sum_{j=1}^{J}\{[\alpha_j(\ln t_{ij}-(\beta_j-\tau_i))]^2\}$$

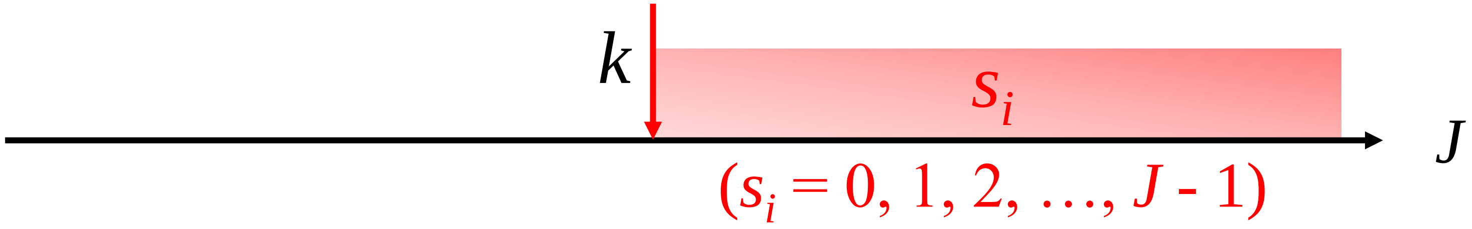
Method

- CPA for item response time data: the likelihood ratio test

$$f(t_{ij};\tau_i,\alpha_j,\beta_j)=\frac{\alpha_j}{t_{ij}\sqrt{2\pi}}exp\{-\frac{1}{2}[\alpha_j(\ln t_{ij}-(\beta_j-\tau_i))]^2\}$$

$$L(\tau_i;\mathbf{t_i})=\prod_{j=1}^{J}\frac{\alpha_j}{t_{ij}\sqrt{2\pi}}exp\{-\frac{1}{2}[\alpha_j(\ln t_{ij}-(\beta_j-\tau_i))]^2\}$$

$$l(\tau_i;\mathbf{t_i})=lnL(\tau_i;\mathbf{t_i})=\sum_{j=1}^{J}\ln\frac{\alpha_j}{t_{ij}\sqrt{2\pi}}-\frac{1}{2}\sum_{j=1}^{J}\{[\alpha_j(\ln t_{ij}-(\beta_j-\tau_i))]^2\}$$



$$\Delta l_i^{(k)} = -2(l_i^{H_0} - l_i^{(k)})$$

Method

- CPA for item response time data: the likelihood ratio test

$$\Delta l_i^{(k)} = -2(l_i^{H_0} - l_i^{(k)})$$

$$l_i^{H_0} = l(\hat{\tau}_{i,0}; \mathbf{t_i})$$

$$l_i^{(k)} = l(\hat{\tau}_{i,k-}; \mathbf{t_{i(k-)}}) + l(\hat{\tau}_{i,k+}; \mathbf{t_{i(k+)}})$$

Method

- CPA for item response time data: the likelihood ratio test

$$\Delta l_i^{(k)} = -2(l_i^{H_0} - l_i^{(k)})$$

$$l_i^{H_0} = l(\hat{\tau}_{i,0}; \mathbf{t_i})$$

$$l_i^{(k)} = l(\hat{\tau}_{i,k-}; \mathbf{t_{i(k-)}}) + l(\hat{\tau}_{i,k+}; \mathbf{t_{i(k+)}})$$

the first k items the remaining items

the change should be in a certain direction: $\tau_{i,k+} > \tau_{i,k-}$

Method

- CPA for item response time data: the likelihood ratio test

$$\Delta l_i^{(k)} = -2(l_i^{H_0} - l_i^{(k)})$$

$$l_i^{H_0} = l(\hat{\tau}_{i,0}; \mathbf{t_i})$$

$$l_i^{(k)} = l(\hat{\tau}_{i,k-}; \mathbf{t_{i(k-)}}) + l(\hat{\tau}_{i,k+}; \mathbf{t_{i(k+)}})$$

the first k items   the remaining items

the change should be in a certain direction: $\tau_{i,k+} > \tau_{i,k-}$

  - the test statistic

$$\Delta l_{\max,i} = \max_{k=1,2,\ldots,(J-1)} \Delta l_i^{(k)}$$

$$\hat{s}_i = J - \arg\max_{k=1,2,\ldots,(J-1)} \{l_i^{(k)}\}$$

Method

- CPA for item response time data: the likelihood ratio test
  - construct the null distribution

$$\Delta l_{\max,i} = \max_{k=1,2,\ldots,(J-1)} \Delta l_i^{(k)}$$ ➡ if it is significantly larger than 0

    - ✓ **challenge:** no closed form distribution
    - ✓ **solution:** ?

Method

- CPA for item response time data: the likelihood ratio test
  - construct the null distribution

$$\Delta l_{\max,i} = \max_{k=1,2,\ldots,(J-1)} \Delta l_i^{(k)}$$ ➡ if it is significantly larger than 0

✓ **challenge:** no closed form distribution

✓ **solution:** ?

Andrews (1993) and Sinharay (2016)

1. the asymptotic critical values

if the change point does not appear too early or too late

TABLE I

ASYMPTOTIC CRITICAL VALUES

		p = 1			p = 2			p = 3			p = 4			p = 5		
π_0	λ	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
.50	1.00	2.69	3.79	6.63	4.61	5.97	9.23	6.25	7.82	11.35	7.77	9.51	13.37	9.23	11.07	15.05
.49	1.08	3.39	4.60	7.63	5.48	6.99	10.35	7.29	8.93	12.77	8.90	10.75	14.75	10.45	12.38	16.61
.48	1.17	3.70	4.96	8.07	5.87	7.41	10.90	7.75	9.43	13.31	9.40	11.27	15.37	11.01	12.95	17.26
.47	1.27	3.93	5.24	8.44	6.19	7.75	11.27	8.08	9.81	13.67	9.80	11.67	15.88	11.42	13.41	17.75
.45	1.49	4.30	5.65	8.93	6.66	8.25	11.89	8.62	10.37	14.31	10.41	12.30	16.57	12.04	14.04	18.49
.40	2.25	4.99	6.40	9.81	7.47	9.13	12.91	9.54	11.35	15.39	11.38	13.37	17.61	13.17	15.19	19.63
.35	3.45	5.49	6.97	10.40	8.10	9.80	13.58	10.25	12.08	16.13	12.15	14.16	18.42	13.94	16.02	20.53
.30	5.44	5.93	7.47	10.84	8.62	10.34	14.10	10.82	12.65	16.65	12.76	14.73	18.96	14.60	16.65	21.12
.25	9.00	6.35	7.87	11.28	9.09	10.78	14.61	11.32	13.18	17.13	13.33	15.29	19.47	15.21	17.27	21.71
.20	16.00	6.73	8.28	11.71	9.54	11.26	15.09	11.81	13.66	17.65	13.82	15.84	19.96	15.76	17.78	22.21
.15	32.11	7.12	8.68	12.16	10.00	11.72	15.56	12.28	14.13	18.07	14.34	16.36	20.47	16.30	18.32	22.66
.10	81.00	7.58	9.11	12.59	10.46	12.17	16.09	12.81	14.69	18.59	14.92	16.91	20.97	16.87	18.86	23.21
.05	361.00	8.13	9.71	13.17	11.08	12.80	16.57	13.46	15.36	19.28	15.64	17.54	21.63	17.58	19.57	23.85

Asymptotic critical values for sup Wald, LM, and LR tests for parameters instability (Donald & Andrews, 2003)

Method

- CPA for item response time data: the likelihood ratio test
  - construct the null distribution

$$\Delta l_{\max, i} = \max_{k=1,2,\ldots,(J-1)} \Delta l_i^{(k)}$$ ➡ if it is significantly larger than 0

✓ **challenge:** no closed form distribution

✓ **solution:** ?

Andrews (1993) and Sinharay (2016)

1. the asymptotic critical values

if the change point does not appear too early or too late

➡ feel time pressure toward

☹ the end of the test

TABLE I

ASYMPTOTIC CRITICAL VALUES

		$p=1$			$p=2$			$p=3$			$p=4$			$p=5$		
π_0	λ	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
.50	1.00	2.69	3.79	6.63	4.61	5.97	9.23	6.25	7.82	11.35	7.77	9.51	13.37	9.23	11.07	15.05
.49	1.08	3.39	4.60	7.63	5.48	6.99	10.35	7.29	8.93	12.77	8.90	10.75	14.75	10.45	12.38	16.61
.48	1.17	3.70	4.96	8.07	5.87	7.41	10.90	7.75	9.43	13.31	9.40	11.27	15.37	11.01	12.95	17.26
.47	1.27	3.93	5.24	8.44	6.19	7.75	11.27	8.08	9.81	13.67	9.80	11.67	15.88	11.42	13.41	17.75
.45	1.49	4.30	5.65	8.93	6.66	8.25	11.89	8.62	10.37	14.31	10.41	12.30	16.57	12.04	14.04	18.49
.40	2.25	4.99	6.40	9.81	7.47	9.13	12.91	9.54	11.35	15.39	11.38	13.37	17.61	13.17	15.19	19.63
.35	3.45	5.49	6.97	10.40	8.10	9.80	13.58	10.25	12.08	16.13	12.15	14.16	18.42	13.94	16.02	20.53
.30	5.44	5.93	7.47	10.84	8.62	10.34	14.10	10.82	12.65	16.65	12.76	14.73	18.96	14.60	16.65	21.12
.25	9.00	6.35	7.87	11.28	9.09	10.78	14.61	11.32	13.18	17.13	13.33	15.29	19.47	15.21	17.27	21.71
.20	16.00	6.73	8.28	11.71	9.54	11.26	15.09	11.81	13.66	17.65	13.82	15.84	19.96	15.76	17.78	22.21
.15	32.11	7.12	8.68	12.16	10.00	11.72	15.56	12.28	14.13	18.07	14.34	16.36	20.47	16.30	18.32	22.66
.10	81.00	7.58	9.11	12.59	10.46	12.17	16.09	12.81	14.69	18.59	14.92	16.91	20.97	16.87	18.86	23.21
.05	361.00	8.13	9.71	13.17	11.08	12.80	16.57	13.46	15.36	19.28	15.64	17.54	21.63	17.58	19.57	23.85

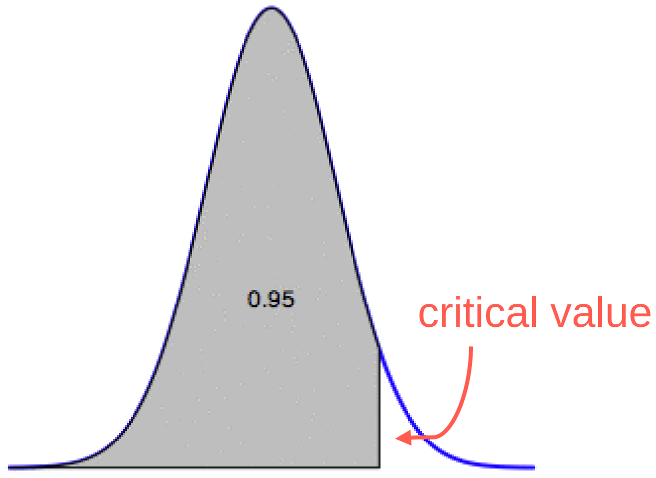
Asymptotic critical values for sup Wald, LM, and LR tests for parameters instability (Donald & Andrews, 2003)

Method

- CPA for item response time data: the likelihood ratio test
  - construct the null distribution

$$\Delta l_{\max,i} = \max_{k=1,2,\ldots,(J-1)} \Delta l_i^{(k)}$$ ➡ if it is significantly larger than 0

- ✓ **challenge:** no closed form distribution
- ✓ **solution:** Monte Carlo simulations



Method

- CPA for item response time data: the likelihood ratio test
  - construct the null distribution

$$\Delta l_{\max,i} = \max_{k=1,2,\ldots,(J-1)} \Delta l_i^{(k)}$$

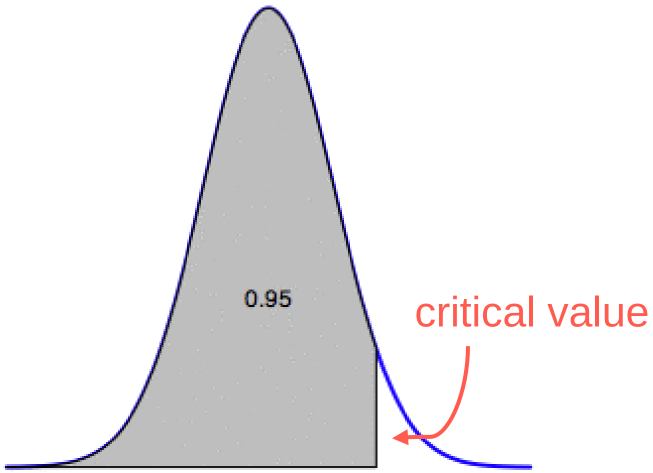
➡ if it is significantly larger than 0

✓ **challenge:** no closed form distribution

✓ **solution:** Monte Carlo simulations

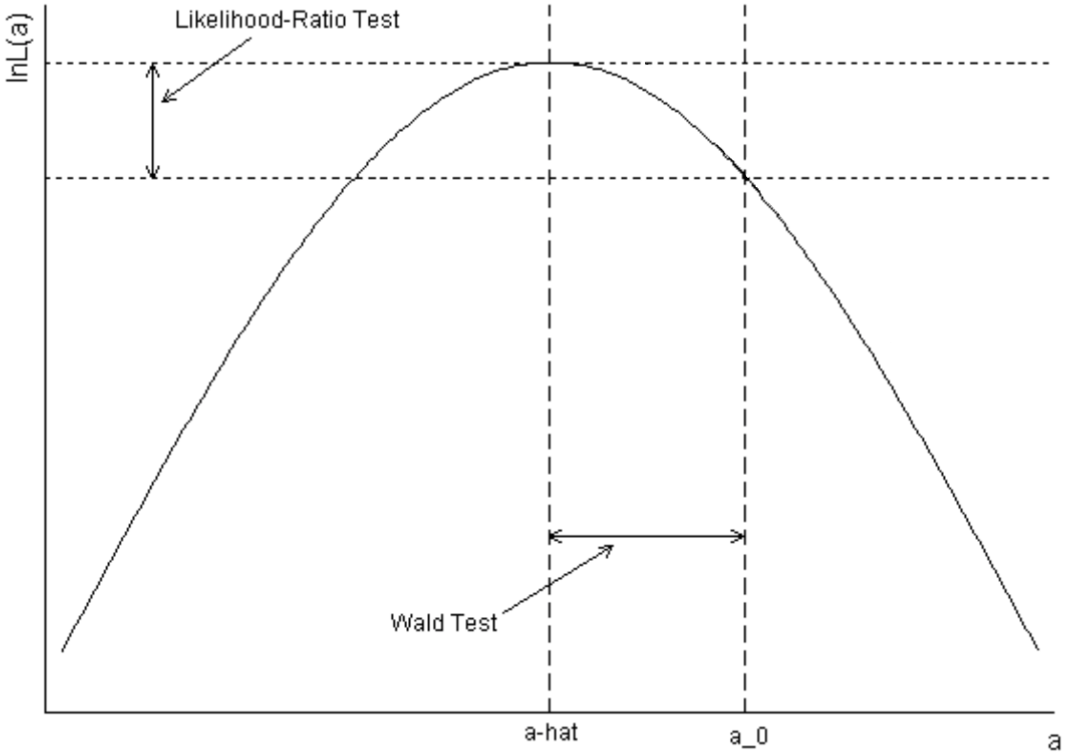
➡ simulating data with no change point to derive the null distribution

1. generate 9,999 values under the null condition
2. take the 1,000th, 500th, and 100th largest values
3. 10%, 5%, and 1% of nominal type-I error level for a one-sided test

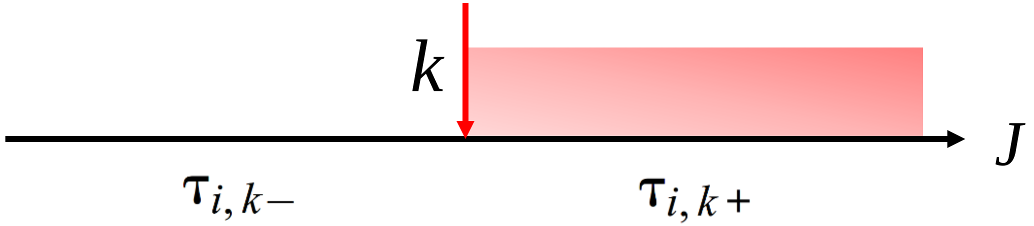


Method

- CPA for item response time data: the Wald test
  - whether the working speed remains unchanged: $\tau_{i,k-} = \tau_{i,k+}$

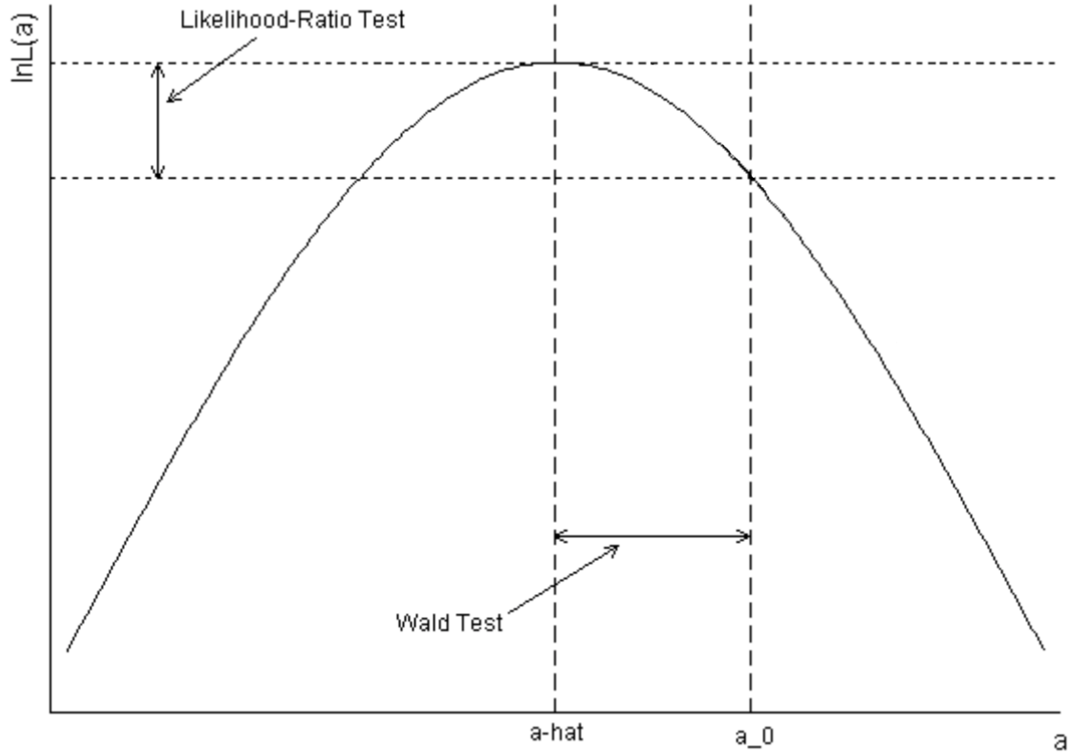


From UCLA, *Statistical Methods and Data Analytics*

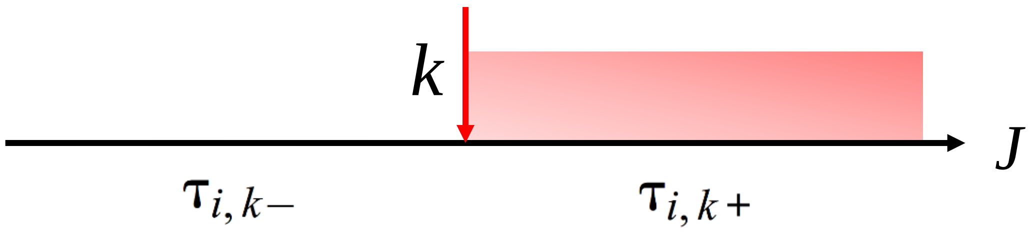


Method

- CPA for item response time data: the Wald test
  - whether the working speed remains unchanged: $\tau_{i,k-} = \tau_{i,k+}$



From UCLA, *Statistical Methods and Data Analytics*



- the test statistic

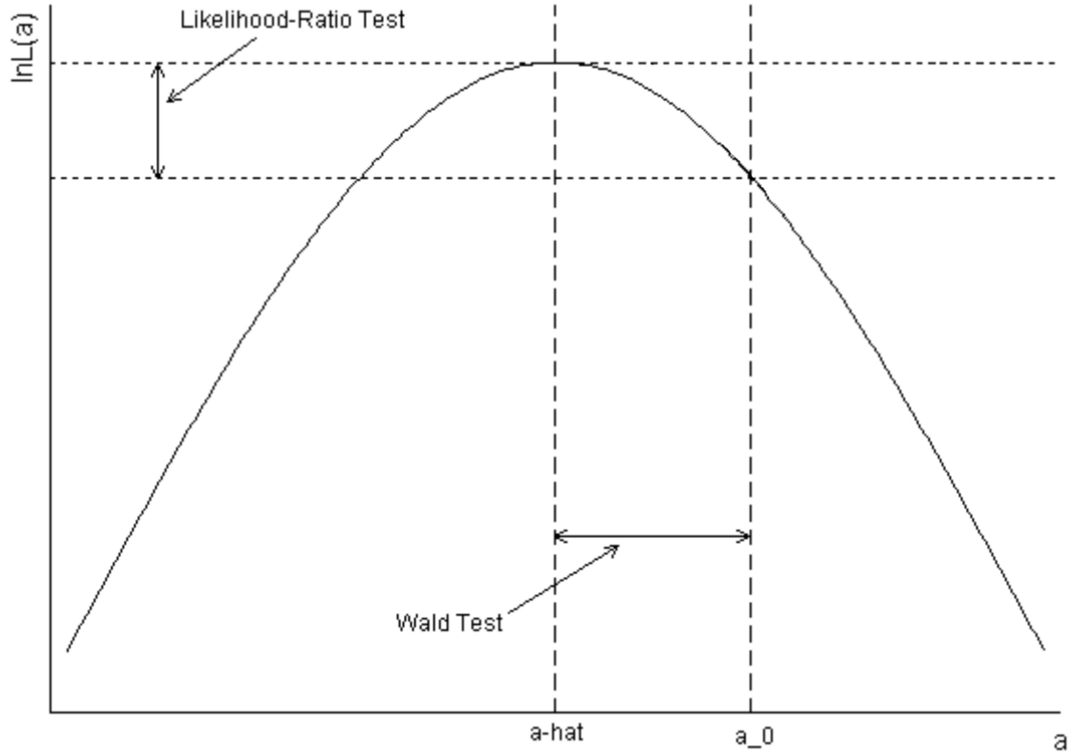
$$W_i^{(k)} = \frac{(\hat{\tau}_{i,k+} - \hat{\tau}_{i,k-})^2}{\frac{1}{I_{k-}(\hat{\tau}_{i,0})} + \frac{1}{I_{k+}(\hat{\tau}_{i,0})}}$$

$$W_{\max,i} = \max_{k=1,2,\ldots,(J-1)} W_i^{(k)}$$

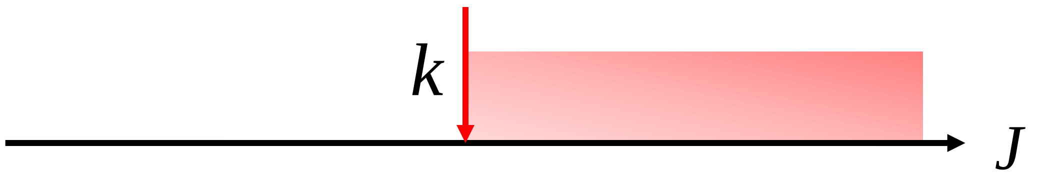
- the critical values: MC simulations

Method

- CPA for item response time data: the Wald test
  - whether the working speed remains unchanged: $\tau_{i,k-} = \tau_{i,k+}$



From UCLA, *Statistical Methods and Data Analytics*



$\tau_{i,k-}$ $\tau_{i,k+}$

- the test statistic

$$W_i^{(k)} = \frac{(\hat{\tau}_{i,k+} - \hat{\tau}_{i,k-})^2}{\frac{1}{I_{k-}(\hat{\tau}_{i,0})} + \frac{1}{I_{k+}(\hat{\tau}_{i,0})}}$$

MLE based on items from 1 to *J*

$$W_{\max,i} = \max_{k=1,2,\ldots,(J-1)} W_i^{(k)}$$

- the critical values: MC simulations

Simulation Study

- **Purpose:** generate the null distribution & evaluate the proposed CPA method
  - the power (TRUE: speeded → speeded)
  - false positive rate/type-I error (TRUE: non-speeded → speeded)

Simulation Study

- **Purpose:** generate the null distribution & evaluate the proposed CPA method
  - the power (TRUE: speeded → speeded)
  - false positive rate/type-I error (TRUE: non-speeded → speeded)

- Response time model:
  - regular patterns: the log-normal model
  - speeded patterns: a gradual change model

Simulation Study

- **Purpose:** generate the null distribution & evaluate the proposed CPA method
  - the power (TRUE: speeded → speeded)
  - false positive rate/type-I error (TRUE: non-speeded → speeded)

- Response time model:
  - regular patterns: the log-normal model
  - speeded patterns: a gradual change model

  Goegebeur et al. (2008):

$$P_{ij}^{*} = \frac{\exp[a_j(\theta_i - b_j)]}{1 + \exp[a_j(\theta_i - b_j)]} * min(1, [1 - (\frac{j}{J} - \eta_i)])^{\lambda_i} \qquad (0 \leq \eta_i \leq 1)$$

Simulation Study

- **Purpose:** generate the null distribution & evaluate the proposed CPA method
  - the power (TRUE: speeded → speeded)
  - false positive rate/type-I error (TRUE: non-speeded → speeded)

- Response time model:
  - regular patterns: the log-normal model
  - speeded patterns: a gradual change model

Goegebeur et al. (2008):

the stage of the test at which examinee i starts to speed
e.g., η_i = .8 → the last 20% of the test

$$P^*_{ij} = \frac{\exp[a_j(\theta_i - b_j)]}{1 + \exp[a_j(\theta_i - b_j)]} * min(1, [1 - (\frac{j}{J} - \eta_i)])^{\lambda_i} \quad (0 \leq \eta_i \leq 1)$$

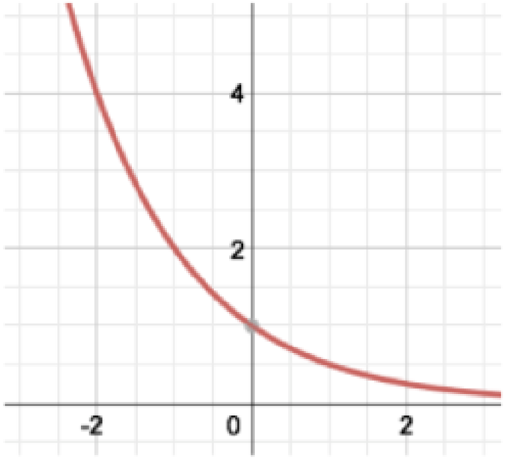
regulate how fast P_{ij}* drops

Simulation Study

- **Purpose:** generate the null distribution & evaluate the proposed CPA method
  - the power (TRUE: speeded → speeded)
  - false positive rate/type-I error (TRUE: non-speeded → speeded)

- Response time model:
  - regular patterns: the log-normal model
  - speeded patterns: a gradual change model

e.g. $b = \frac{1}{2} \Rightarrow f(x) = \left(\frac{1}{2}\right)^x$



Goegebeur et al. (2008):

$$P^*_{ij} = \frac{\exp[a_j(\theta_i - b_j)]}{1 + \exp[a_j(\theta_i - b_j)]} * min(1, [1 - (\frac{j}{J} - \eta_i)])^{\lambda_i}$$

when the test does not reach the speeded stage: 1 (2PLM)

when the test reaches the speeded stage: $[1 - (\frac{j}{J} - \eta_i)]^{\lambda_i}$

Simulation Study

- **Purpose:** generate the null distribution & evaluate the proposed CPA method
  - the power (TRUE: speeded → speeded)
  - false positive rate/type-I error (TRUE: non-speeded → speeded)

- Response time model:
  - regular patterns: the log-normal model
  - speeded patterns: a gradual change model

Goegebeur et al. (2008):

$$P_{ij}^{*} = \frac{\exp[a_j(\theta_i - b_j)]}{1 + \exp[a_j(\theta_i - b_j)]} * min(1, [1 - (\frac{j}{J} - \eta_i)])^{\lambda_i}$$

Similarly:

$$\ln(t_{ij}) = (\beta_j - \tau_i + \varepsilon_{ij}) * min(1, [1 - (\frac{j}{J} - \eta_i)])^{\lambda_i}, \ \varepsilon_{ij} \sim N(0, \alpha_j^{-2})$$

Simulation Study

- **Purpose:** generate the null distribution & evaluate the proposed CPA method
  - the power (TRUE: speeded → speeded)
  - false positive rate/type-I error (TRUE: non-speeded → speeded)

- Response time model:
  - regular patterns: the log-normal model
  - speeded patterns: a gradual change model

Goegebeur et al. (2008):

$$P_{ij}^{*} = \frac{\exp[a_j(\theta_i - b_j)]}{1 + \exp[a_j(\theta_i - b_j)]} * min(1, [1 - (\frac{j}{J} - \eta_i)])^{\lambda_i}$$

Similarly:

$$\ln(t_{ij}) = (\beta_j - \tau_i + \varepsilon_{ij}) * min(1, [1 - (\frac{j}{J} - \eta_i)])^{\lambda_i}, \ \ \varepsilon_{ij} \sim N(0, \alpha_j^{-2})$$

✓ **be more realistic & allow to evaluate the robustness**

✓ **regulate the change point**

✓ **keep parallel to Shao et al. (2016)**

Simulation Study

- Simulation Design
  - test length: 40, 60, and 80 items
  - time limit: 60, 90, and 120 minutes, respectively
  - sample size: N = 1,000
  - the percentage of speeded test takers: 10%, 30%

  - run out of time: unreached (response time = 0)
    (the test taker would be labeled as ‘‘speeded’’)

Simulation Study

- Simulation Design
  - speeded parameters:

    $\lambda \sim logN(3.912, 1)$

    η : beta distribution (median, $\eta = .6$ or $.7$ and $\eta_{var} = .001$ or $40 \times .001 = .04$)

    the change point: $40 \times 0.6 = 25$

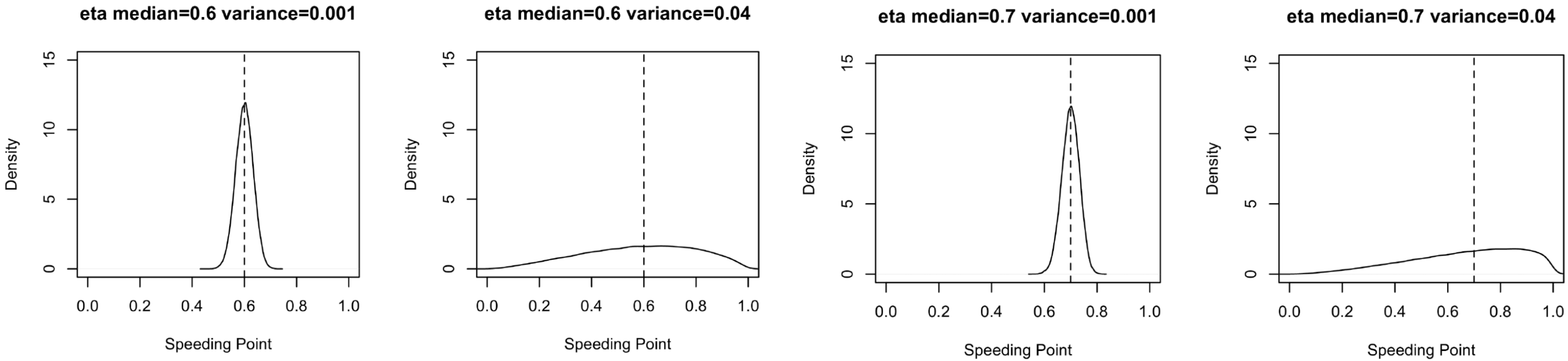
Simulation Study

- Simulation Design
  - speeded parameters:

    $\lambda \sim logN(3.912, 1)$

    η : beta distribution (median, η = .6 or .7 and η_{var} = .001 or $40 \times .001 = .04$)

    the change point: $40 \times 0.6 = 25$



Simulation Study

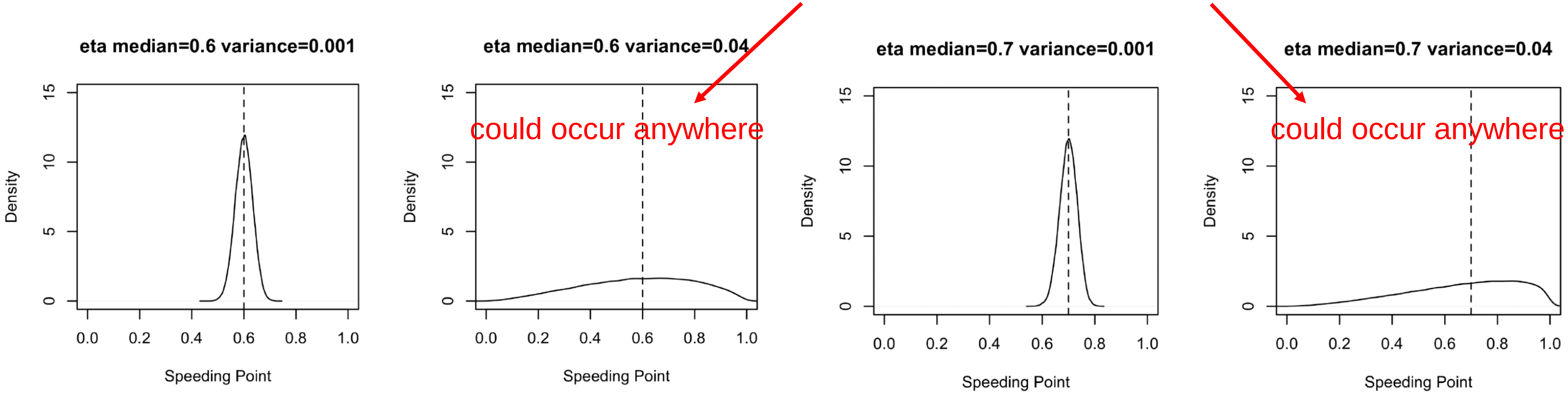
- Simulation Design
  - speeded parameters:

    $\lambda \sim logN(3.912, 1)$

    η : beta distribution (median, η = .6 or .7 and η_{var} = .001 or $40 \times .001 = .04$)

    the change point: 40 × 0.6 = 25

evaluate the validity the asymptotic critical values (late or middle)



Simulation Study

- Simulation Design
  - response time parameters:

    $\tau \sim N(0, 0.25)$

    $\alpha \sim U(1.75, 3.25)$

    β: mean = 4, SD = 1/3, $\rho_{\beta a} = 0.3$, $\rho_{\beta b} = 0.5$

  - response item parameters:

    $a \sim lnN(0, 0.5)$, $b \sim N(0, 1)$

    random normal deviates were added to a linear combination of a and b

    to produce $\rho_{\beta a}$ and $\rho_{\beta b}$

Simulation Study

- Simulation Design
  - response time parameters:

    $\tau \sim N(0, 0.25)$

    $\alpha \sim U(1.75, 3.25)$

    β: mean = 4, SD = 1/3, $\rho_{a\beta} = 0.3$, $\rho_{b\beta} = 0.5$

  - response item parameters:

    $a \sim lnN(0, 0.5)$, $b \sim N(0, 1)$

    random normal deviates were added to a linear combination of a and b

    to produce $\rho_{\beta a}$ and $\rho_{\beta b}$

→

a total of 24 conditions = 3 (test length) * 2 (percentage) * 2 (η_{mean}) * 2 ($\eta_{variance}$)

each condition was replicated 50 times

Simulation Study

- Find the critical values
  - simulating 10,000 no speeded response time patterns

    → 10,000 test statistics $\Delta l_{\max,i}$ and $W_{max,i}$

    → choose the 500th (0.05), 100th (0.01), and 10th (0.001) largest values

    → each of the null condition was replicated 1,000 times

    → take the average as the empirical critical values

Simulation Study

- Find the critical values
  - simulating 10,000 no speeded response time patterns

    → 10,000 test statistics $\Delta l_{\max,i}$ and $W_{max,i}$

    → choose the 500th (0.05), 100th (0.01), and 10th (0.001) largest values

    → each of the null condition was replicated 1,000 times

    → take the average as the empirical critical values

almost identical

Table 1. Mean (and *SD*) of the Critical Values for Wald and Likelihood Ratio Test.

J	$c_{.05}$	$c_{.01}$	$c_{.001}$
40	8.148 (0.09)	11.345 (0.19)	15.772 (0.59)
60	8.293 (0.09)	11.483 (0.20)	15.883 (0.64)
80	8.765 (0.09)	12.024 (0.20)	16.533 (0.64)

Simulation Study

- Find the critical values

Sinharay

TABLE 1.
Asymptotic Critical Values for the Distribution of the Supremum of the Square of a Standardized Tied-Down Bessel Process

	Significance Level		
$100 \times n_1/n$	1%	5%	10%
5	13.01	9.84	8.19
10	12.69	9.31	7.63
15	12.35	8.85	7.17
20	11.69	8.45	6.80

Table 1. Mean (and *SD*) of the Critical Values for Wald and Likelihood Ratio Test.

J	$c_{.05}$	$c_{.01}$	$c_{.001}$
40	8.148 (0.09)	11.345 (0.19)	15.772 (0.59)
60	8.293 (0.09)	11.483 (0.20)	15.883 (0.64)
80	8.765 (0.09)	12.024 (0.20)	16.533 (0.64)

Simulation Study

- Find the critical values

restricted conditions:
1. long test length
2. change point should occur in the middle of the test

Sinharay

TABLE 1.
Asymptotic Critical Values for the Distribution of the Supremum of the Square of a Standardized Tied-Down Bessel Process

	Significance Level		
$100 \times n_1/n$	1%	5%	10%
5	13.01	9.84	8.19
10	12.69	9.31	7.63
15	12.35	8.85	7.17
20	11.69	8.45	6.80

Table 1. Mean (and *SD*) of the Critical Values for Wald and Likelihood Ratio Test.

J	$c_{.05}$	$c_{.01}$	$c_{.001}$
40	8.148 (0.09)	11.345 (0.19)	15.772 (0.59)
60	8.293 (0.09)	11.483 (0.20)	15.883 (0.64)
80	8.765 (0.09)	12.024 (0.20)	16.533 (0.64)

Simulation Study

- Find the critical values

    1. select $\alpha_{0.05} = 8$ and $\alpha_{0.01} = 11$

Table 1. Mean (and *SD*) of the Critical Values for Wald and Likelihood Ratio Test.

J	$c_{.05}$	$c_{.01}$	$c_{.001}$
40	8.148 (0.09)	11.345 (0.19)	15.772 (0.59)
60	8.293 (0.09)	11.483 (0.20)	15.883 (0.64)
80	8.765 (0.09)	12.024 (0.20)	16.533 (0.64)

Simulation Study

- Find the critical values

  1. select $\alpha_{0.05} = 8$ and $\alpha_{0.01} = 11$
  2. also included $\alpha_{0.05} = 8.85$ and $\alpha_{0.01} = 12.35$
     $\rightarrow$ to compare with Andrews' (1993)

TABLE 1

ASYMPTOTIC CRITICAL VALUES
FOR TESTS OF PARAMETER INSTABILITY WITH $\Pi = [.15, .85]^{a}$

	Significance Level			
Degrees of Freedom (p_0)	1%	2.5%	5%	10%
1	12.3	10.3	8.7	7.2
2	15.3	13.4	11.7	10.1
3	18.3	16.0	14.2	12.3
4	20.7	18.2	16.3	14.4
5	22.6	20.4	18.4	16.4

Table 1. Mean (and *SD*) of the Critical Values for Wald and Likelihood Ratio Test.

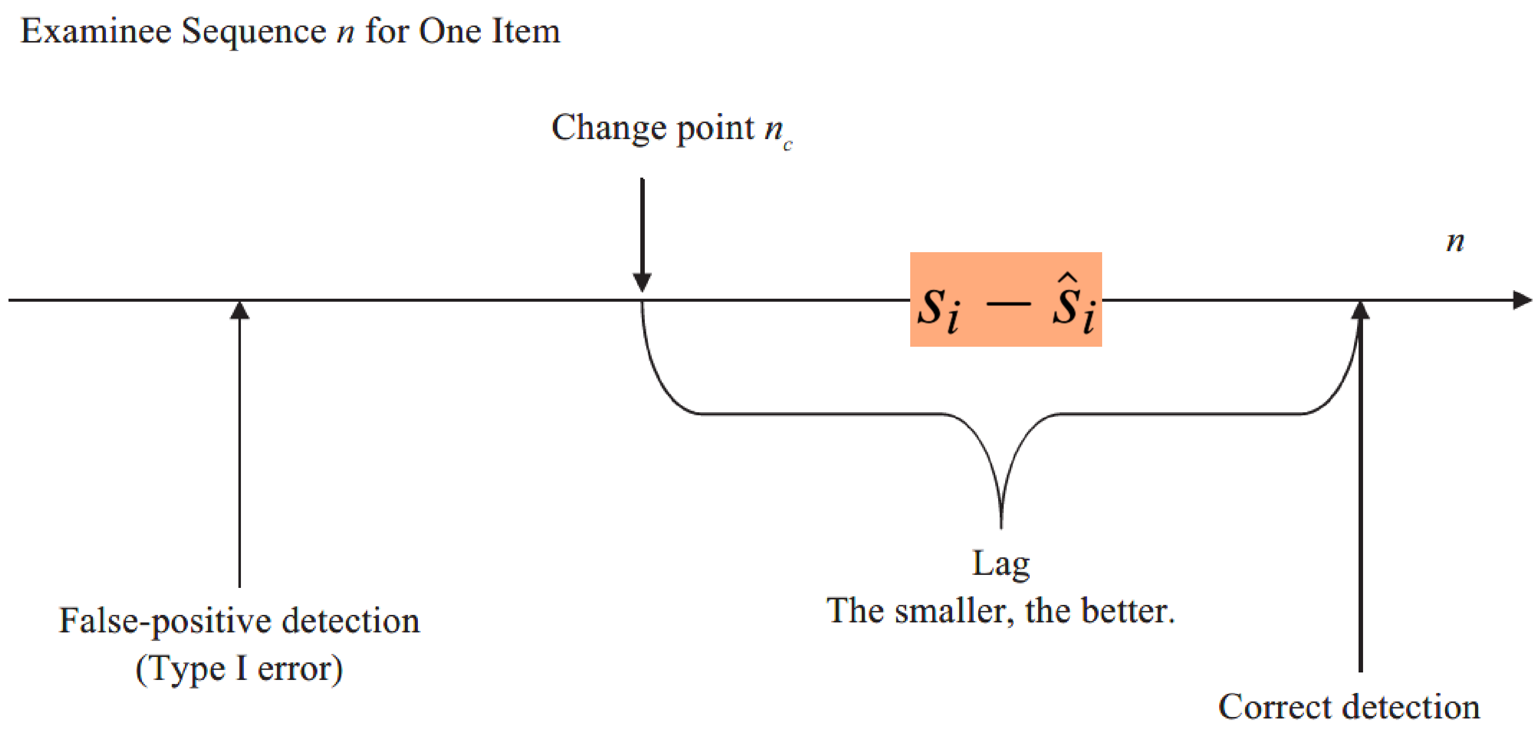
J	$c_{.05}$	$c_{.01}$	$c_{.001}$
40	8.148 (0.09)	11.345 (0.19)	15.772 (0.59)
60	8.293 (0.09)	11.483 (0.20)	15.883 (0.64)
80	8.765 (0.09)	12.024 (0.20)	16.533 (0.64)

Results

- Two points of interest:
  - examine the empirical **power** and **false positive rate**
  - evaluate the performance of estimating the actual change point: **the lag** (≈ bias)

Results

- Two points of interest:
  - examine the empirical **power** and **false positive rate**
  - evaluate the performance of estimating the actual change point: **the lag** (≈ bias)



Zhang, 2014 *APM*

Results

- Two points of interest:
  - examine the empirical **power** and **false positive rate**
  - evaluate the performance of estimating the actual change point: **the lag** (bias)

    the standard error (RMSE)

    the average of the absolute value: AL_{mean}

    the absolute value of the lag divided by the test length: $AL_{mean\%}$

    ➡ to make the results **across different test lengths** comparable

Results

Table 2. Power, False Positive Rate, %*NF*, (Absolute) Lag, Bias, and RMSE for $J = 40$.

			Power				False positive rate								
%	η_{median}	η_{var}	$C_{.05}$	$A_{.05}$	$C_{.01}$	$A_{.01}$	$C_{.05}$	$A_{.05}$	$C_{.01}$	$A_{.01}$	% *NF*	AL_{mean}	$AL_{mean\%}$	Bias	RMSE
0	—	—	—	—	—	—	0.053	0.035	0.012	0.006	—	—	—	—	—
10	.6	.001	1.00	1.00	1.00	1.00	0.054	0.035	0.012	0.006	4.79	1.17	0.03	1.08	1.77
10	.7	.001	1.00	1.00	1.00	1.00	0.055	0.037	0.012	0.006	4.75	1.17	0.03	1.01	1.58
10	.6	.04	1.00	1.00	1.00	1.00	0.056	0.036	0.012	0.006	4.91	2.25	0.06	1.05	3.22
10	.7	.04	0.98	0.98	0.98	0.97	0.053	0.035	0.011	0.006	5.04	3.76	0.09	0.40	6.49
30	.6	.001	1.00	1.00	1.00	1.00	0.053	0.034	0.012	0.006	3.92	1.16	0.03	1.07	1.68
30	.7	.001	1.00	1.00	1.00	1.00	0.051	0.032	0.011	0.006	3.86	1.32	0.03	1.01	1.82
30	.6	.04	0.99	0.99	0.98	0.98	0.056	0.036	0.012	0.005	3.99	6.10	0.15	0.81	8.63
30	.7	.04	0.94	0.94	0.93	0.92	0.056	0.037	0.013	0.007	3.98	6.61	0.17	−0.14	9.69

$C_{.05}$ = applying the proposed simple critical value at $\alpha_{.05}$; $A_{.05}$ = applying Andrews' critical value at $\alpha_{.05}$; $C_{.01}$ and $A_{.01}$ = similar to $C_{.05}$ and $A_{.05}$ but at $\alpha_{.01}$; % *NF* = percentage of test takers that did not finish the test in time; AL_{mean} = the mean absolute lag between the actual and estimated change point across replications; $AL_{mean\%}$ = the mean absolute lag between the actual and estimated change point divided by the test length across replications; Bias = the bias of change point estimate; RMSE = the root mean square error of the change point estimate.

Results

Table 3. Power, False Positive Rate, % *NF*, (Absolute) Lag, Bias, and RMSE for J = 60.

			Power				False positive rate								
%	η_{median}	η_{var}	$C_{.05}$	$A_{.05}$	$C_{.01}$	$A_{.01}$	$C_{.05}$	$A_{.05}$	$C_{.01}$	$A_{.01}$	% *NF*	AL_{mean}	$AL_{mean\%}$	Bias	RMSE
0	—	—	—	—	—	—	0.057	0.037	0.013	0.007	—	—	—	—	—
10	.6	.001	1.00	1.00	1.00	1.00	0.059	0.038	0.014	0.007	7.00	1.65	0.03	1.49	2.47
10	.7	.001	1.00	1.00	1.00	1.00	0.057	0.036	0.013	0.006	6.96	1.66	0.03	1.47	2.37
10	.6	.04	0.98	0.98	0.98	0.98	0.057	0.038	0.013	0.007	7.28	6.25	0.10	1.20	9.37
10	.7	.04	0.98	0.98	0.98	0.98	0.056	0.037	0.012	0.006	7.21	6.81	0.11	1.28	10.16
30	.6	.001	1.00	1.00	1.00	1.00	0.056	0.037	0.012	0.006	5.37	1.57	0.03	1.48	2.28
30	.7	.001	1.00	1.00	1.00	1.00	0.059	0.038	0.012	0.006	5.50	1.97	0.03	1.55	2.70
30	.6	.04	0.99	0.99	0.99	0.99	0.058	0.038	0.013	0.006	5.53	8.83	0.15	1.17	12.74
30	.7	.04	0.97	0.96	0.96	0.96	0.059	0.038	0.013	0.007	5.99	11.63	0.19	0.27	16.01

Results

Table 4. Power, False Positive Rate, % *NF*, (Absolute) Lag, Bias and RMSE for J = 80.

			Power				False positive rate								
%	η_{median}	η_{var}	$C_{.05}$	$A_{.05}$	$C_{.01}$	$A_{.01}$	$C_{.05}$	$A_{.05}$	$C_{.01}$	$A_{.01}$	% *NF*	AL_{mean}	$AL_{mean\%}$	Bias	RMSE
0	—	—	—	—	—	—	0.073	0.049	0.017	0.008	—	—	—	—	—
10	.6	.001	1.00	1.00	1.00	1.00	0.071	0.046	0.016	0.008	6.43	2.31	0.03	2.26	3.32
10	.7	.001	1.00	1.00	1.00	1.00	0.073	0.049	0.017	0.008	6.45	2.34	0.03	1.99	3.12
10	.6	.04	0.99	0.99	0.99	0.99	0.072	0.048	0.017	0.009	6.53	6.53	0.08	1.85	9.54
10	.7	.04	0.98	0.98	0.97	0.96	0.071	0.047	0.017	0.009	6.67	10.71	0.13	0.88	16.89
30	.6	.001	1.00	1.00	1.00	1.00	0.072	0.047	0.015	0.008	5.02	2.12	0.03	1.93	2.98
30	.7	.001	1.00	1.00	1.00	1.00	0.073	0.048	0.017	0.009	5.22	2.71	0.03	2.07	3.53
30	.6	.04	0.99	0.99	0.99	0.99	0.070	0.046	0.016	0.009	5.51	15.17	0.19	1.66	20.82
30	.7	.04	0.97	0.97	0.96	0.96	0.073	0.048	0.018	0.009	5.52	16.91	0.21	0.65	22.72

Results

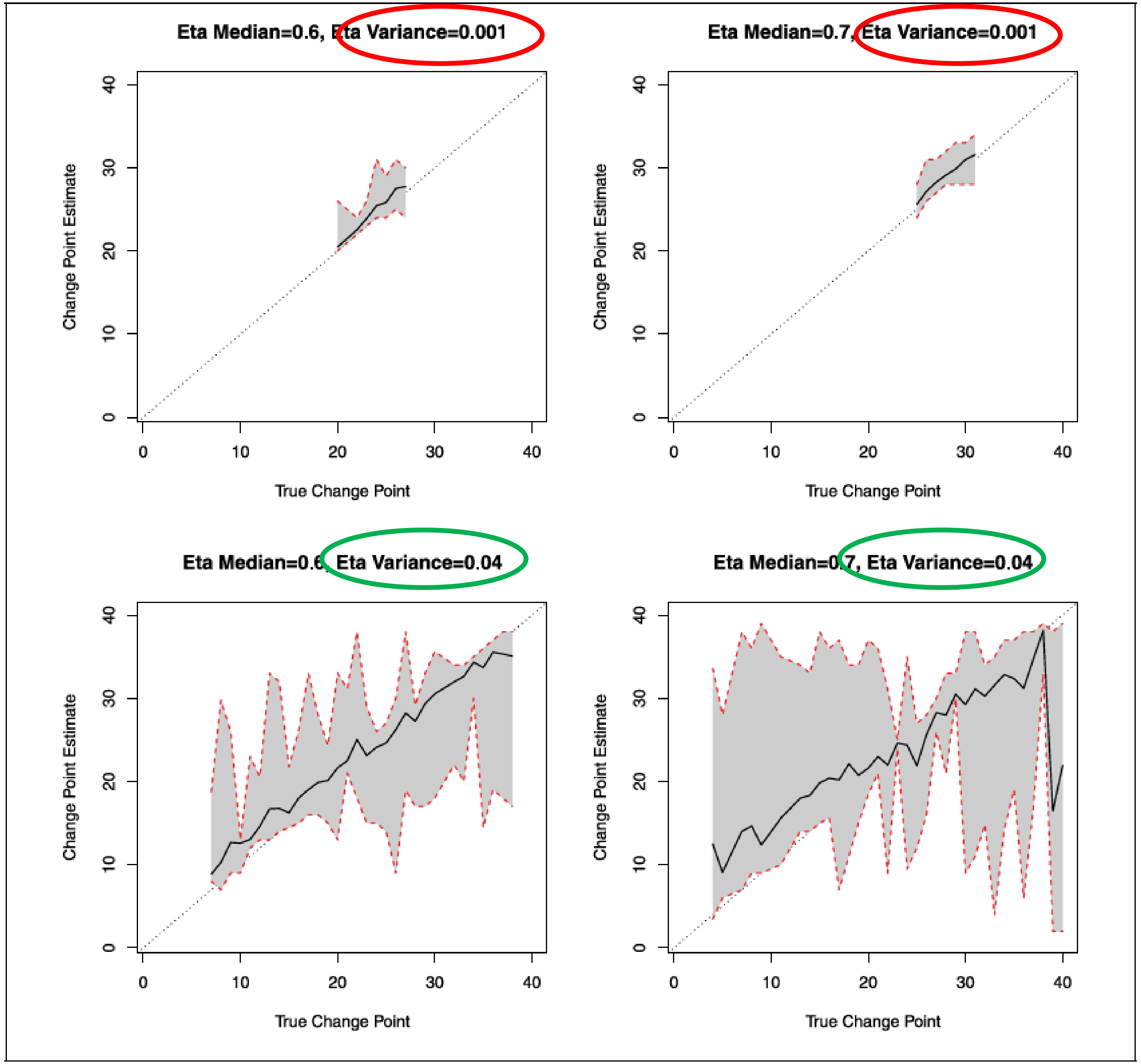


Figure 2. Estimated versus true change point with 10% of speeded test takers at test length of 40.

Results

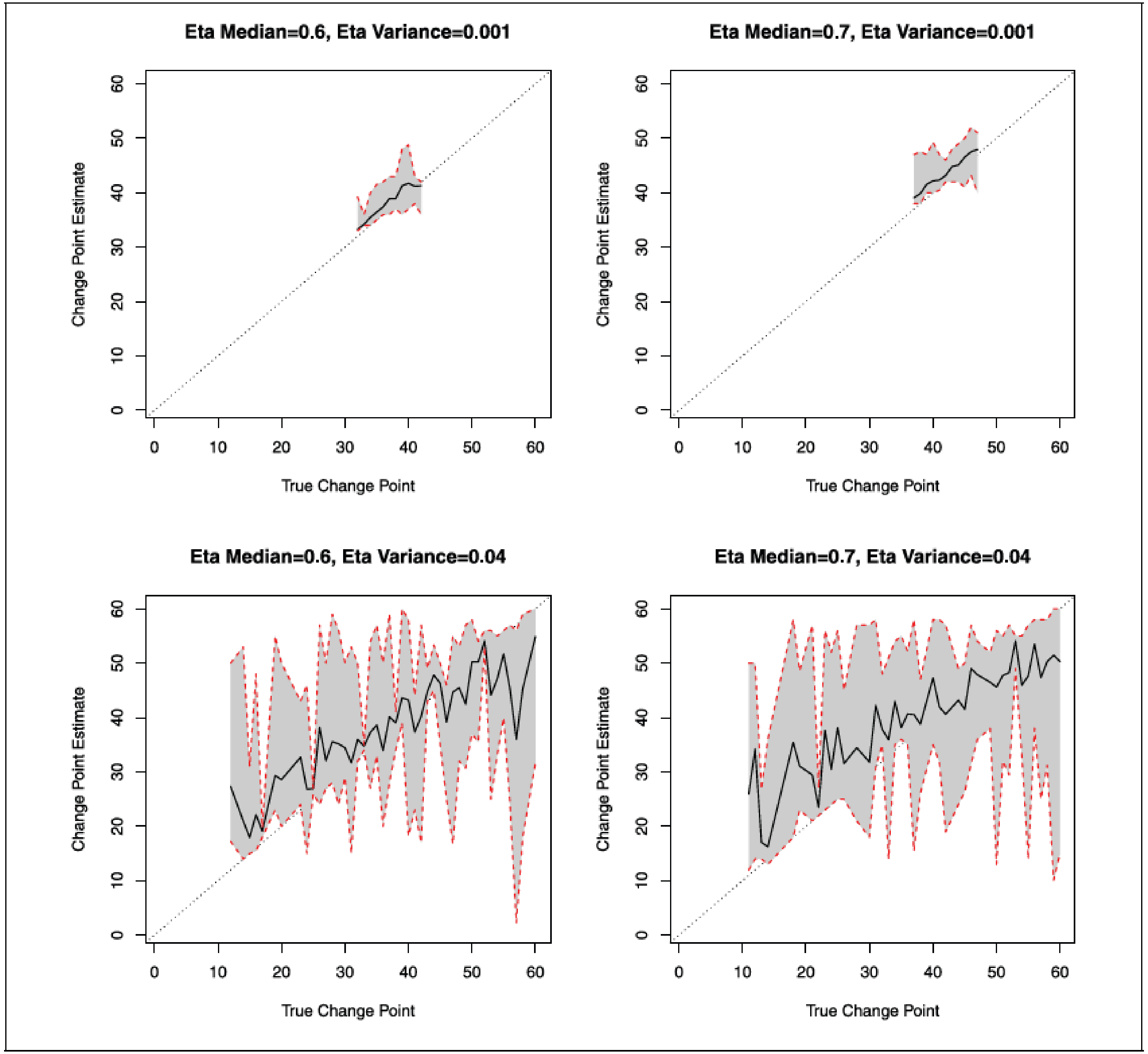


Figure 3. Estimated versus true change point with 10% of speeded test takers at test length of 60.

Results

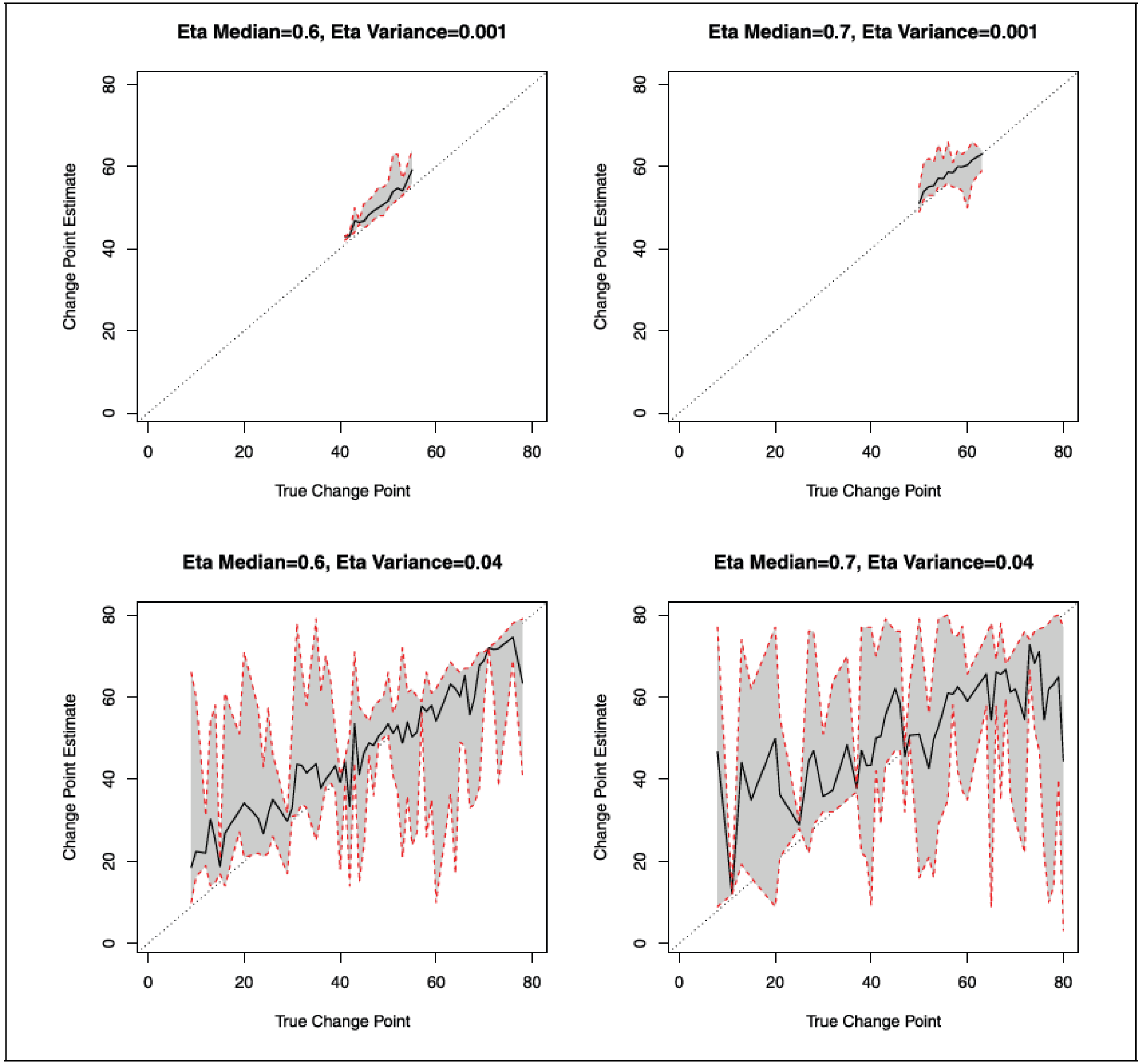


Figure 4. Estimated versus true change point with 10% of speeded test takers at test length of 80.

Real Data Analysis

- the application of the proposed CPA method
  - 50,000 test takers on a 30-item multiple-choice computer-based assessment
  - remove test retakers and those who finished the entire test within 5 minutes
  - remove cases with response time of 0 on late items (conspicuous cases)

  - around 46,000 test takers
    (split into five samples each containing data from 9,200 students)

    ➡ cross-validate the findings

Real Data Analysis

- the application of the proposed CPA method
  - 50,000 test takers on a 30-item multiple-choice computer-based assessment
  - remove test retakers and those who finished the entire test within 5 minutes
  - remove cases with response time of 0 on late items (conspicuous cases)

  - around 46,000 test takers

    (split into five samples each containing data from 9,200 students)

    ➡ cross-validate the findings

  - fit the log-normal model
  - obtain α_j and β_j (treated as known and unchanging when estimate τ)

    ➡ $\Delta l_{\max,i}$ $W_{max,i}$ + Shao et al. (2016)'s: Δl_i (100 permutation)

Results

- the computational gains (for each sample)

hardware specification:

6 core, 2.90 GHz Intel Core i5-9400 processor, and 16.0 GB RAM

  - using 100 random permutation: 5 hours

  - using the simple cutoffs: less than 1 minute
    maximum memory used around 400M

Results

Table 5. Number of Test Takers Flagged and the Percent of Flagged Using Response Time and Item Response.

	$\alpha_{.05}$			$\alpha_{.01}$		
	Response time			Response time		
Sample	Current	Andrews	Response	Current	Andrews	Response
1	1384 (15.0%)	1152 (12.5%)	276 (3.0%)	812 (8.8%)	687 (7.5%)	41 (0.4%)
2	1442 (15.7%)	1205 (13.1%)	348 (3.8%)	835 (9.1%)	683 (7.4%)	71 (0.8%)
3	1450 (15.8%)	1214 (13.2%)	296 (3.2%)	837 (9.1%)	715 (7.8%)	58 (0.6%)
4	1427 (15.5%)	1203 (13.1%)	300 (3.3%)	831 (9.0 %)	702 (7.6%)	57 (0.6%)
5	1385 (15.1%)	1182 (12.8%)	270 (2.9%)	828 (9.0%)	688 (7.5%)	53 (0.6%)

Results

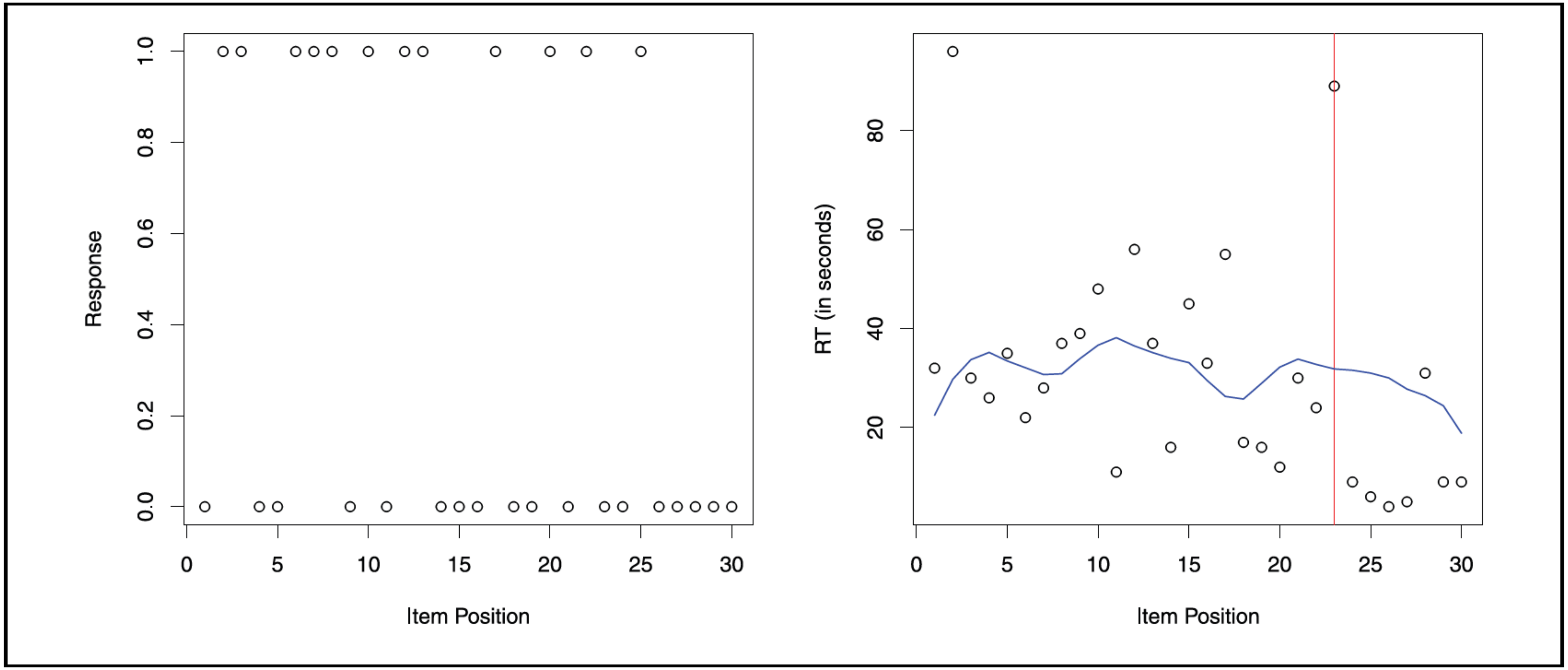


Figure 5. The response and response time pattern of test taker (first example).

Results

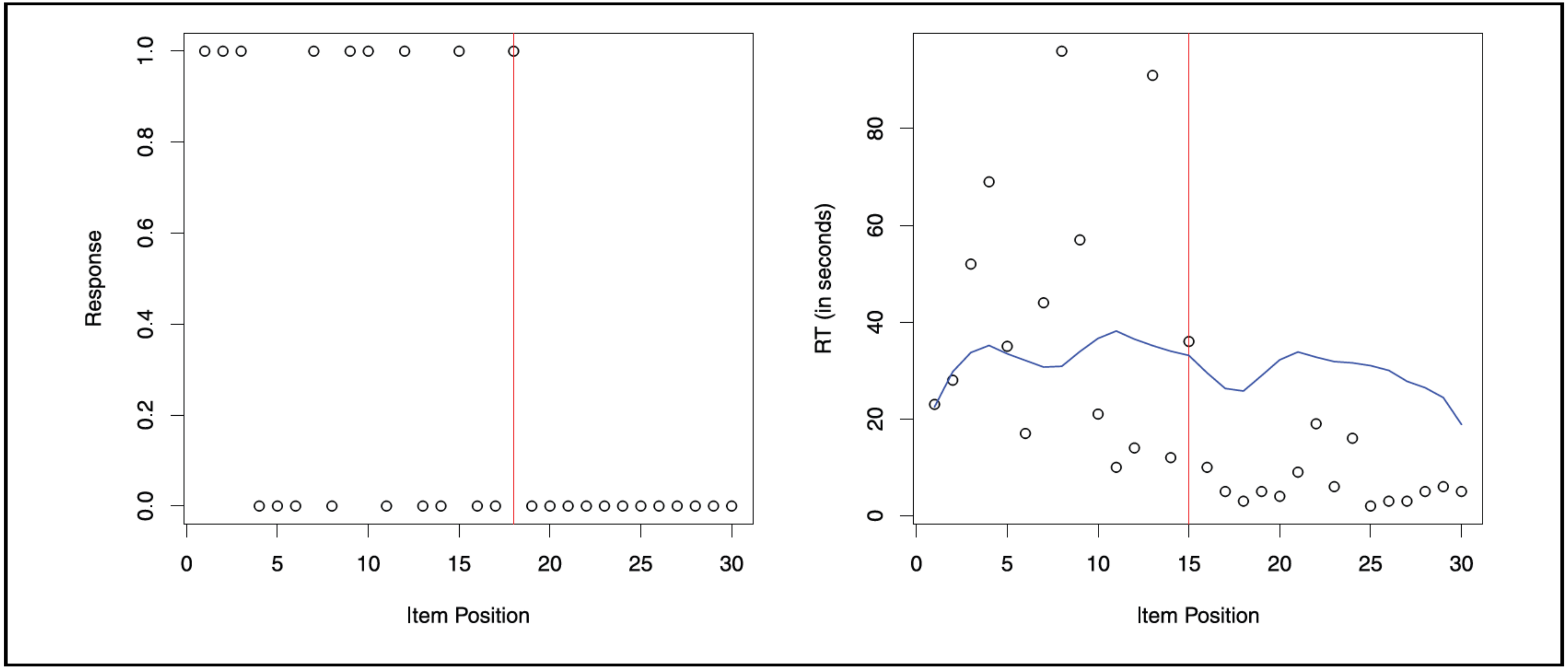


Figure 6. The response and response time pattern of test taker (second example).

Conclusions and Discussion

- propose a CPA method to detect test speededness using response time
- success: show high power in detecting speeded examinees, even when simple and fixed critical values are used
- flexible:

1. can also be applied to other types of response time models
2. other types of aberrant responses (fatigue and inattentiveness)
3. the simple setting of critical values
4. tests with polytomous items, as well as mixed-format tests
5. be applicable to CAT or a multi-stage testing (stable with different lengths)

Limitations

- different assumptions:

  responses: speededness manifests itself in performance decline

  response times: speededness will manifest in faster responding

  (both should be considered & both are challenged)

  → the assumption will be valid in what context (e.g., high-stakes testing) and to what extent?

  → using the hierarchical model

- explore model-free quickest change detection methods
- pinpoint the cause of change point (e.g., time limit or low motivation)

Practical implications

- important quality control components
- understand the prevalence of speededness

- *One should exercise extreme caution when it comes to removing any response or test taker data, and one should refrain from relying solely on statistical results to make such decisions.*

- *Typically human review should follow statistical quality control procedures, and it should be no different when they are applied to testing.*

Thanks for Listening!

Reporter: Yingshi Huang