PSYCHOMETRIKA—VOL. 83, NO. 3, 650–673
SEPTEMBER 2018
https://doi.org/10.1007/s11336-017-9596-3

IF(5 YEARS): 2.510

Psychometric Society

CrossMark

SEQUENTIAL DETECTION OF COMPROMISED ITEMS USING RESPONSE TIMES IN COMPUTERIZED ADAPTIVE TESTING

Edison M. Choe
Graduate Management
Admission Council

Jinming Zhang
University of Illinois,
Urbana-Champaign

hua-hua Chang
University of Illinois,
Urbana-Champaign

Reporter: Yingshi Huang

Introduction

- Computerized Adaptive Testing (CAT)
- **A GLARING SECURITY ISSUE**
  - items are sequentially selected
  - maximize information selection method: highly **unbalanced item exposure**

➡ ① Sympson–Hetter (SH) method
② a-stratification techniques
③ …

➡ a realistic item pool size << the number of examinees

spot anomalous behavior of both examinees and items!

Introduction

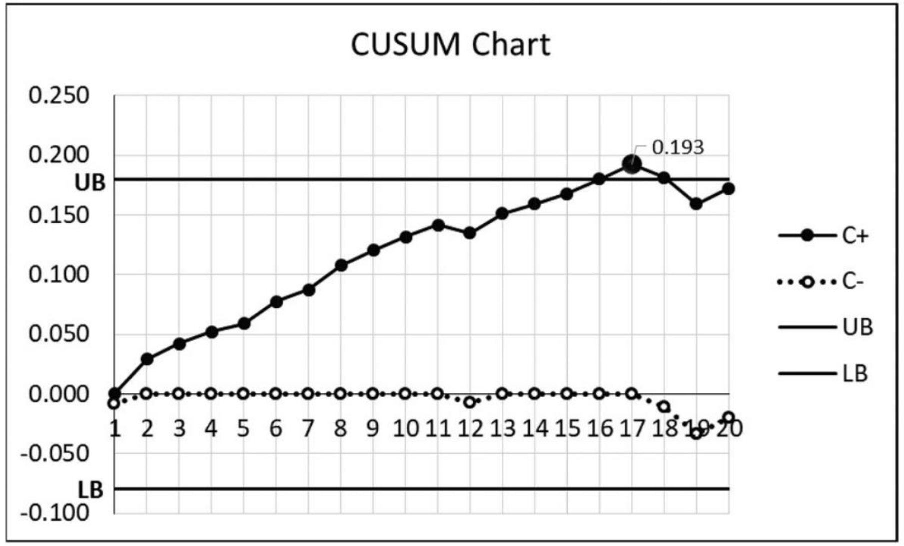
- From the **examinee** perspective

− detect an aberrant pattern of responses or response times (RTs)

- From the **item** perspective √

− detect item parameter drift (IPD)

☹ CUSUM: need to repeat item calibration at each sequential step



[Xiaofeng Yu & Ying Cheng, 2020, figure 1]

➡ **inadequate sample size**

tremendous computational burden

Introduction

- From the **examinee** perspective

− detect an aberrant pattern of responses or response times (RTs)

- From the **item** perspective √

− detect item parameter drift (IPD)

− detect an aberrant pattern of responses or RTs across all examinees that have been administered the item

Belov, 2014

O'Leary & Smith, 2017

McLeod & Schnipke, 1999

need to identify a larger set of potentially aberrant examinees first

Introduction

- From the **examinee** perspective

− detect an aberrant pattern of responses or response times (RTs)

- From the **item** perspective √

− detect item parameter drift (IPD)

− detect an aberrant pattern of responses or RTs across all examinees that have been administered the item

Belov, 2014
O'Leary & Smith, 2017
McLeod & Schnipke, 1999

need to identify a larger set of potentially aberrant examinees first

Lu & Hambleton, 2003
Han & Hambleton, 2004
Zhang, 2014; Zhang & Li, 2016

✓ real-time detection procedure
✓ quick & relatively high accuracy

Δ responses / every exposure

+

Δ RTs / every exposure
PURPOSE

CAT Framework

- Response model
  - $P(X_{ij}=1|\theta)=P_j(\theta_i)=c_j+\dfrac{1-c_j}{1+\mathrm{e}^{-a_j(\theta_i-b_j)}}$
  - $I_j(\theta_i)=-E\left(\dfrac{\partial^2}{\partial\theta_i^2}\log L(\theta_i|x_{ij})\right)=a_j^2\left(\dfrac{1-P_j(\theta_i)}{P_j(\theta_i)}\right)\left(\dfrac{P_j(\theta_i)-c_j}{1-c_j}\right)^2$
  - $\mathrm{SE}\left(\hat{\theta}_i^{ML}\right)\approx\dfrac{1}{\sqrt{I^{(k)}\left(\hat{\theta}_i^{ML}\right)}}=\dfrac{1}{\sqrt{\sum_{j=1}^{k}I_j\left(\hat{\theta}_i^{ML}\right)}}$

unbalanced item pool usage

↓

at greater risk of compromise

How to reduce item exposure?

CAT Framework

- the Sympson–Hetter (SH) method

$p(S)$ → the probability that an item is **‘selected’**

$p(A)$ → the probability that an item is actually **‘administered’**

$$p(A) = p(A|S) \times p(S) \leq r_{\max}$$

to adjust $p(S)$ such that $p(A)$ is less than or equal to r_{max}

a random number is less than $p(A|S)$: administer

otherwise: select next item

CAT Framework

- the Sympson–Hetter (SH) method

$p(S)$ → the probability that an item is **'selected'**

$p(A)$ → the probability that an item is actually **'administered'**

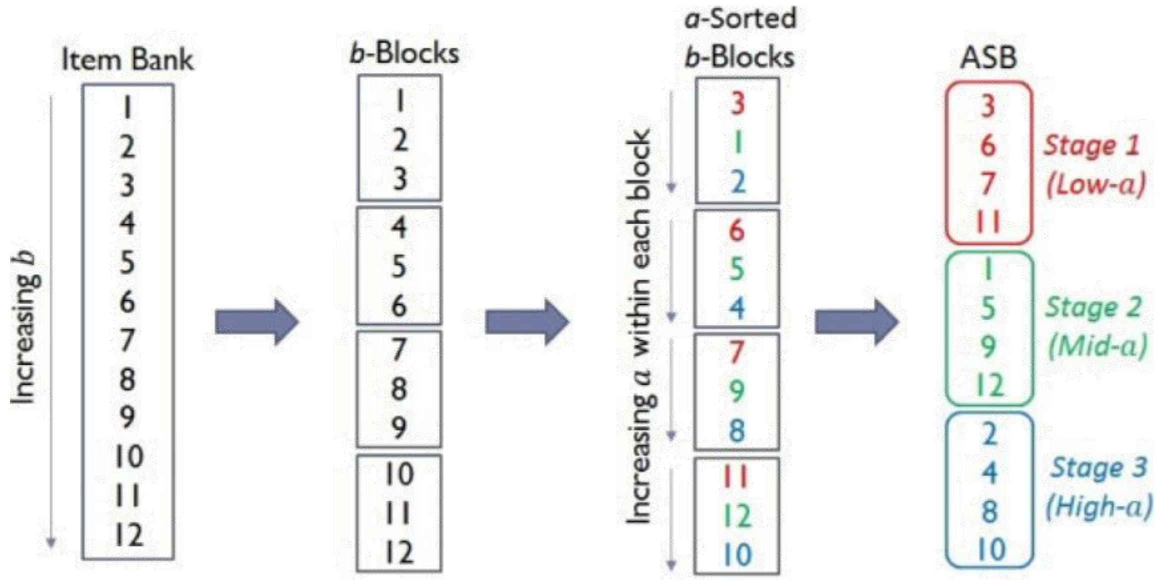
➡ $p(A) = p(A|S) \times p(S) \leq r_{\max}$

unable to increase exposure for underexposed items

√ • a-stratification with b-blocking (ASB)

at any given stage:

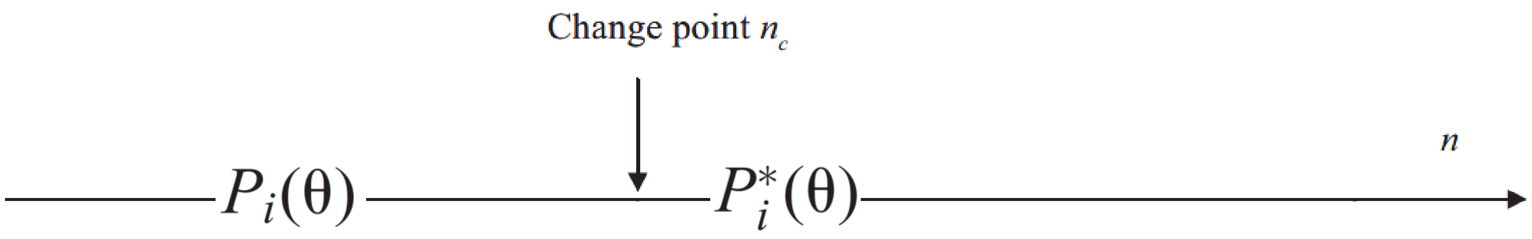
maximize $B_j(\hat{\theta}_i) = \dfrac{1}{|\hat{\theta}_i - b_j|}$



Sequential Monitoring Procedures

- Using Responses (based on IRT)

Examinee Sequence *n* for One Item

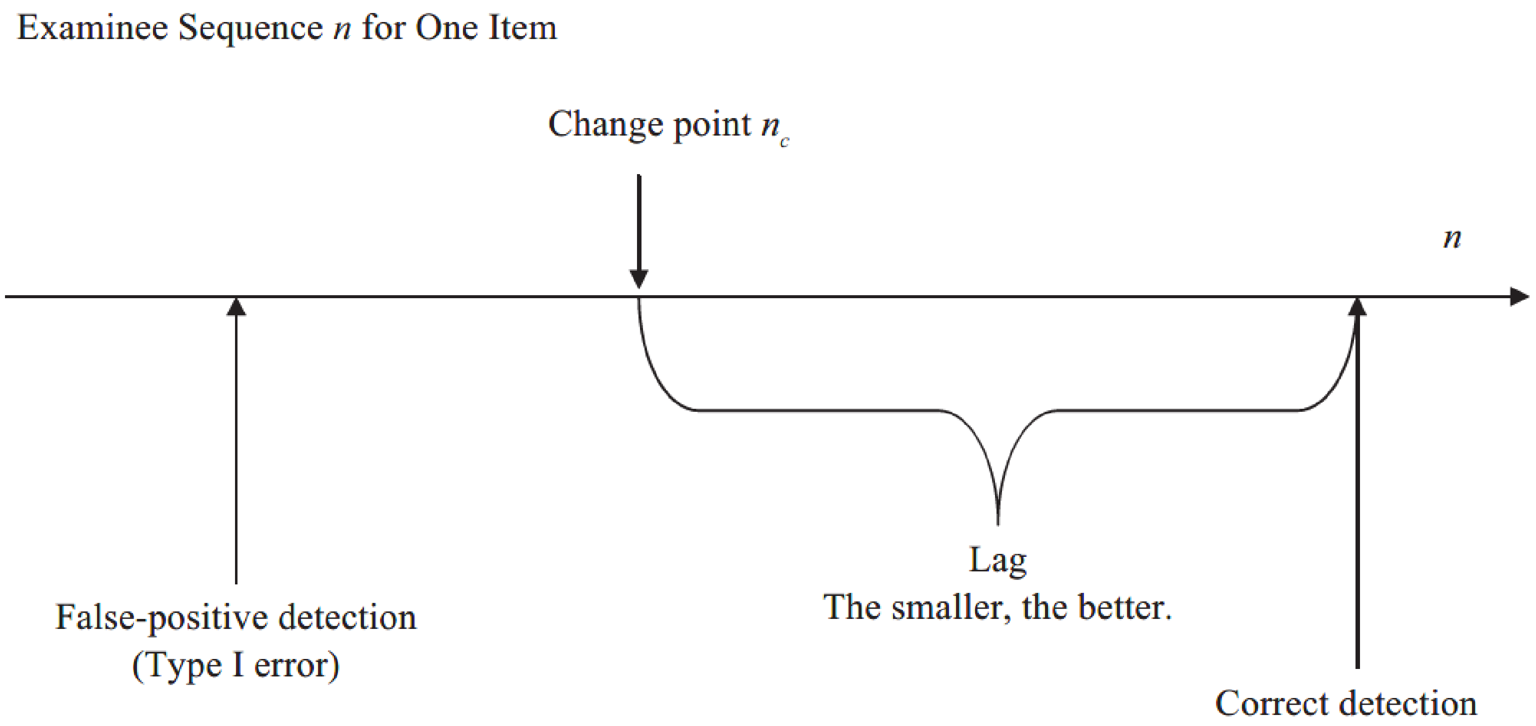


$P_i(\theta) \leq P_i^*(\theta)$ at each θ level

Zhang, 2014 *APM*

Sequential Monitoring Procedures

- Using Responses (based on IRT)



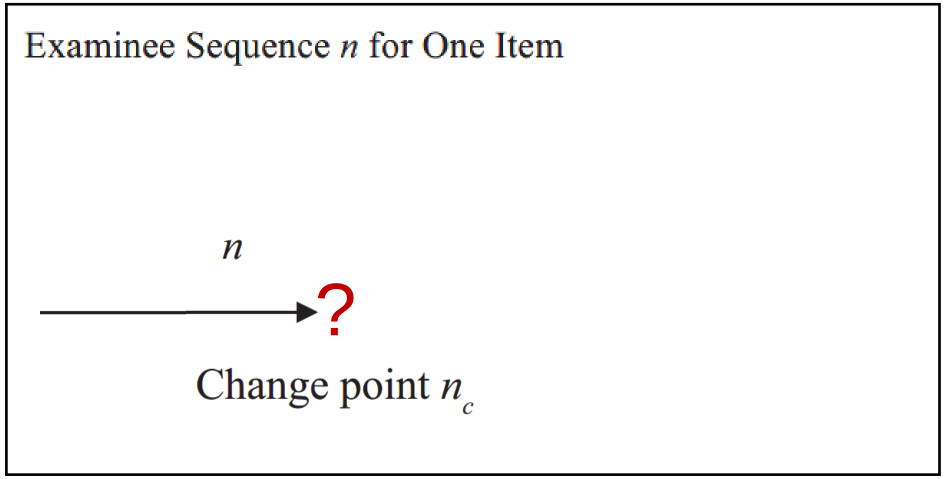
The objective:

✓ detect **significant increase** in the number of correct responses as soon as possible
✓ control the rate of false detections

Zhang, 2014 *APM*

Sequential Monitoring Procedures

- Using Responses (based on IRT)



$$H_0 : \sum^{n} X_{ij} = \sum^{n} P_j(\theta_i)$$

$$H_1 : \sum^{n} X_{ij} > \sum^{n} P_j(\theta_i)$$

  - The observation: $\sum^{n} X_{ij}$
  - The expectation: $\sum^{n} P_j(\theta_i)$

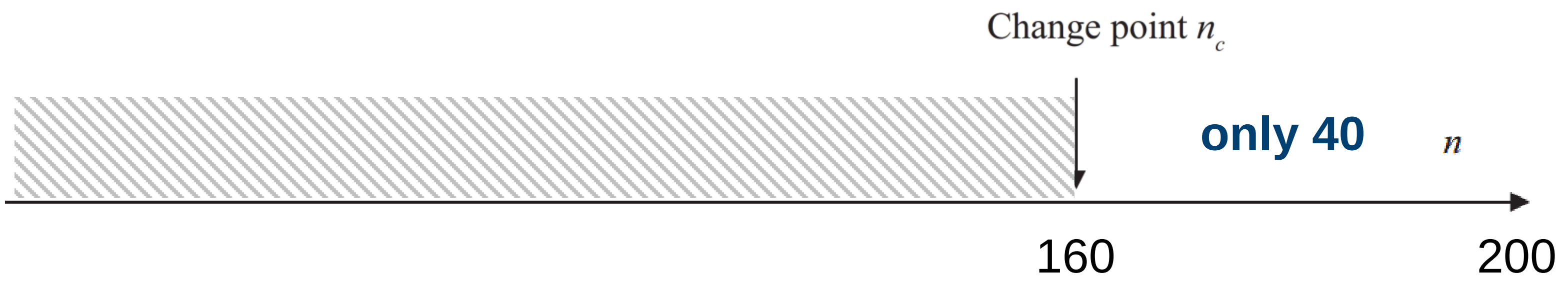
    (benchmark value: when the item is not compromised)

Zhang & Li, 2016 *JEM*

Sequential Monitoring Procedures

- Using Responses (based on IRT)

Examinee Sequence n for One Item



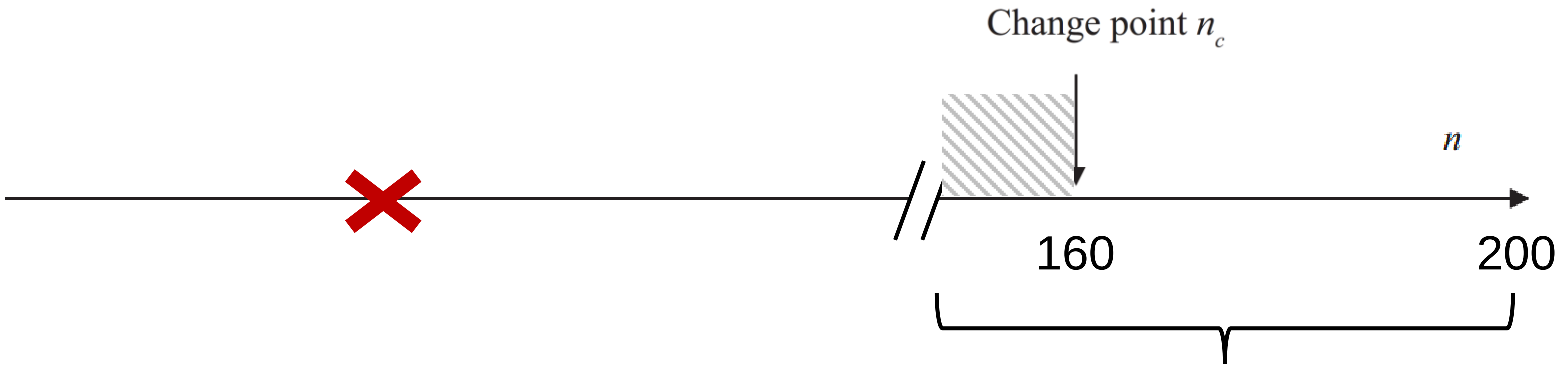
160:40

not sensitive to the change

moving sample: use the **most recent** responses instead

$1 \rightarrow n$

$n - m + 1 \rightarrow n$



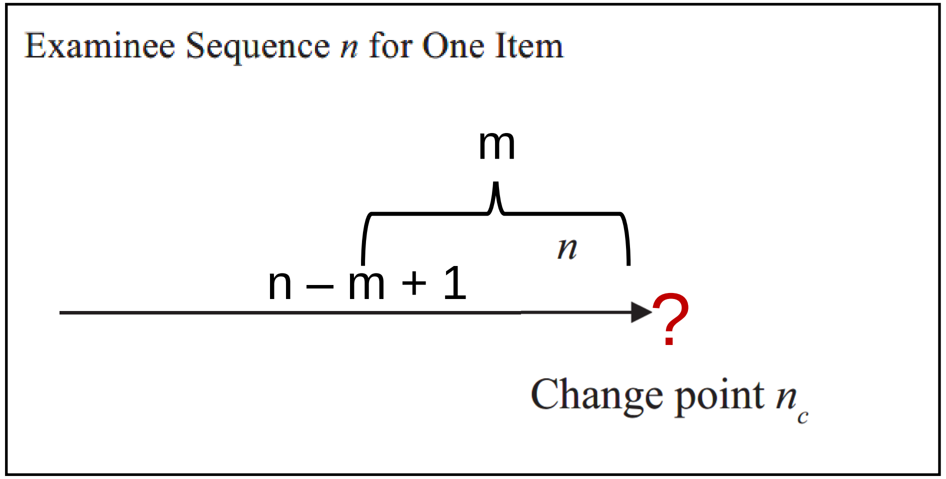
e.g., m = 50

10:40

Zhang, 2014 *APM*

Sequential Monitoring Procedures

- Using Responses (based on IRT)



$$H_0 : p_j^{(m)} = \sum_{i=n-m+1}^{n} P_j(\theta_i)/m$$

$$H_1 : p_j^{(m)} > \sum_{i=n-m+1}^{n} P_j(\theta_i)/m$$

- The observation: $Y_j^{(m)} = \sum_{i=n-m+1}^{n} X_{ij}$ and $\hat{p}_j^{(m)} = Y_j^{(m)}/m$
- The expectation: $E\left(Y_j^{(m)}\right) = \sum_{i=n-m+1}^{n} P_j(\theta_i)$

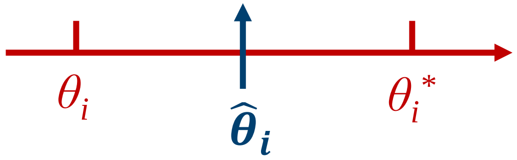
Zhang & Li, 2016 *JEM*

Sequential Monitoring Procedures

- Using Responses (based on IRT)

$$H_0 : p_j^{(m)} = \sum_{i=n-m+1}^{n} P_j(\theta_i)/m$$ → true θ_i is never known

$$H_1 : p_j^{(m)} > \sum_{i=n-m+1}^{n} P_j(\theta_i)/m$$



positively biased → diminish the power

– to construct a test statistic:

➢ X_{ij} is a Bernoulli random variable

$$E(X_{ij}) = P_j(\theta_i), \quad Var(X_{ij}) = P_j(\theta_i)(1 - P_j(\theta_i))$$

➢ $Y_j^{(m)}$ is a Poisson-binomial random variable

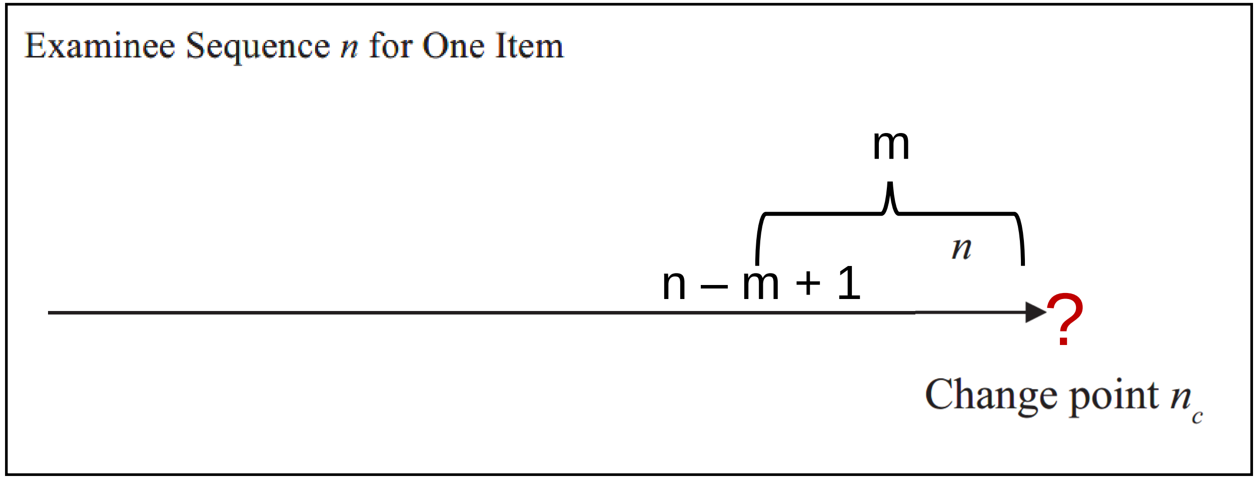
$$E\left(Y_j^{(m)}\right) = \sum_{i=n-m+1}^{n} P_j(\theta_i), \quad Var\left(Y_j^{(m)}\right) = \sum_{i=n-m+1}^{n} P_j(\theta_i)(1 - P_j(\theta_i))$$

$$\Rightarrow \frac{\hat{p}_j^{(m)} - \sum_{i=n-m+1}^{n} P_j(\theta_i)/m}{\sqrt{\sum_{i=n-m+1}^{n} P_j(\theta_i)(1 - P_j(\theta_i))/m^2}} \xrightarrow{d} \mathcal{N}(0, 1) \text{ under } H_0$$

Zhang & Li, 2016 *JEM*

Sequential Monitoring Procedures

- Using Responses (based on CTT)



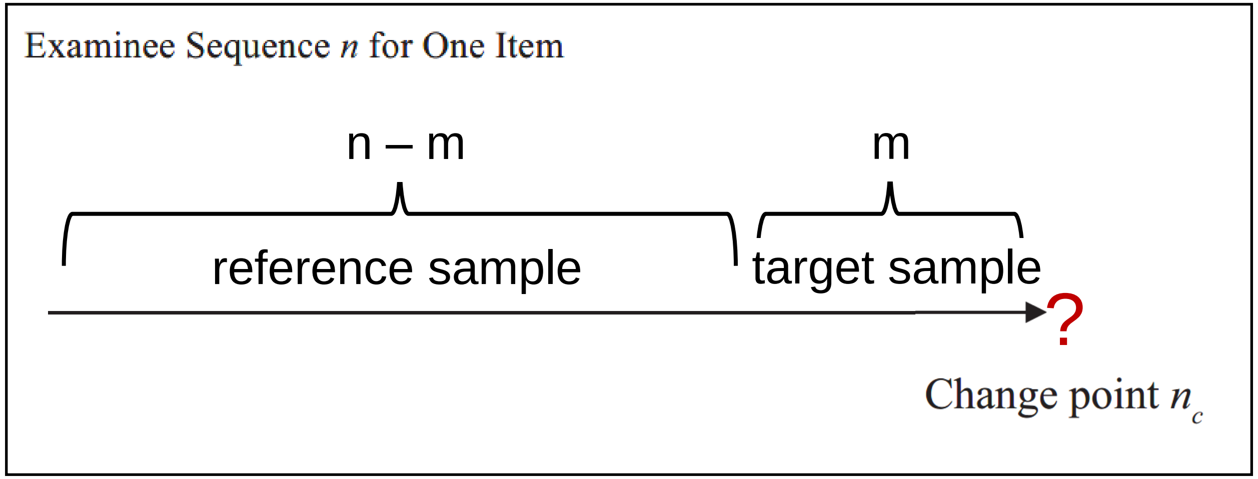
  - The observation: $Y_j^{(m)} = \sum_{i=n-m+1}^{n} X_{ij}$ and $\hat{p}_j^{(m)} = Y_j^{(m)}/m$
  - The expectation: ?

    Find a reference sample (when the item is not compromised)

Zhang, 2014 *APM*

Sequential Monitoring Procedures

- Using Responses (based on CTT)



$$H_0 : p_j^{(m)} = p_j^{(r)}$$

$$H_1 : p_j^{(m)} > p_j^{(r)}$$

- The observation: $Y_j^{(m)} = \sum_{i=n-m+1}^{n} X_{ij}$ and $\hat{p}_j^{(m)} = Y_j^{(m)}/m$
- The expectation: $\hat{p}_j^{(r)} = \frac{\sum_{i=1}^{n-m} X_{ij}}{n-m}$ (empirical benchmark)

Zhang, 2014 *APM*

Sequential Monitoring Procedures

- Using Responses (based on CTT)

$$H_0 : p_j^{(m)} = p_j^{(r)}$$
$$H_1 : p_j^{(m)} > p_j^{(r)}$$

– to construct a test statistic:

$$\mathrm{Var}(\hat{p}_j^{(m)} - \hat{p}_j^{(r)}) = p_j(1 - p_j)\left(\frac{1}{m} + \frac{1}{n-m}\right) \rightarrow \hat{p}_j = \frac{(n-m)\hat{p}_j^{(r)} + m\hat{p}_j^{(m)}}{(n-m)+m} = \frac{\sum_{i=1}^{n} X_{ij}}{n}$$

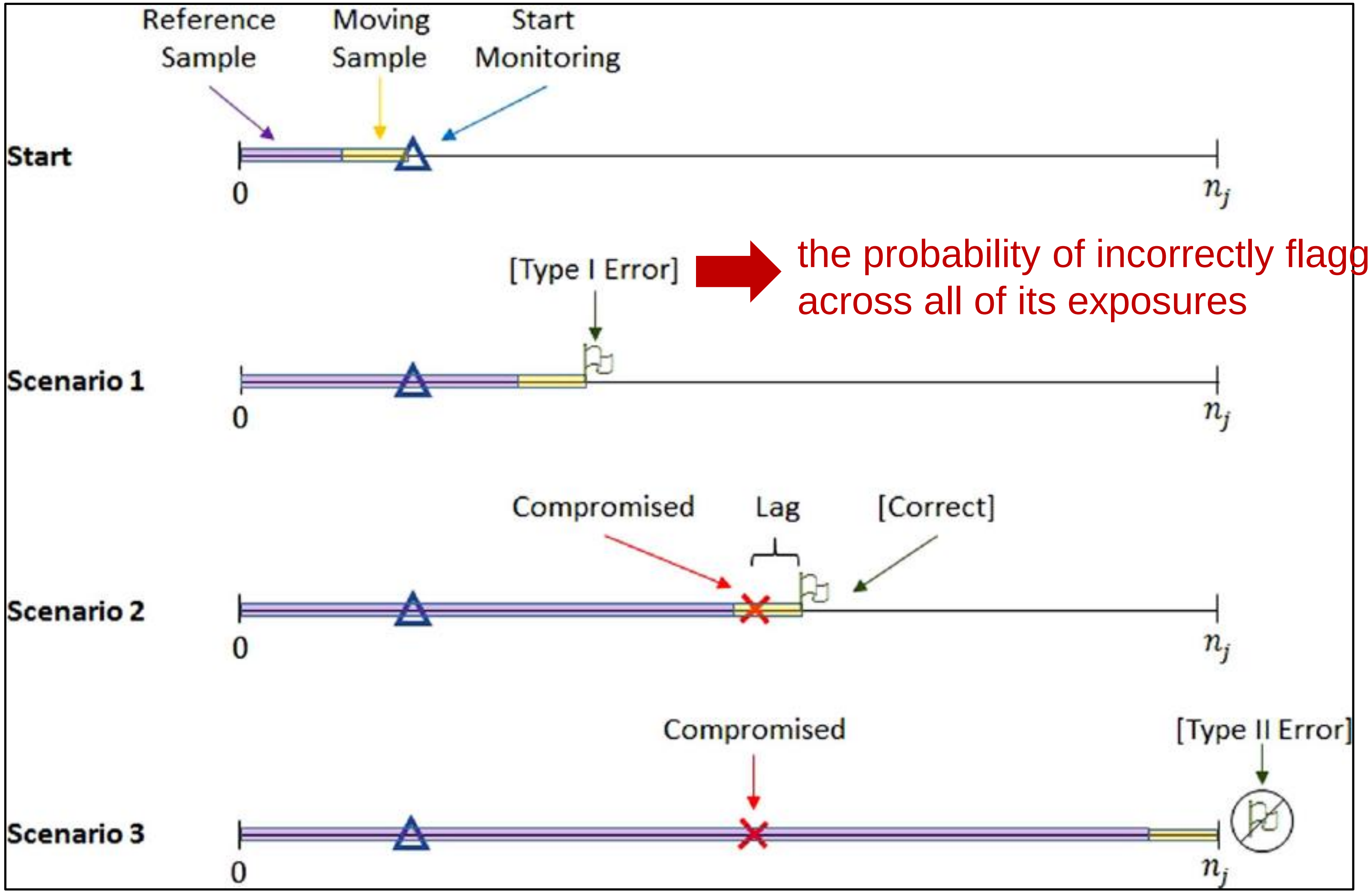
$$Z_j = \frac{\hat{p}_j^{(m)} - \hat{p}_j^{(r)}}{\sqrt{\hat{p}_j(1-\hat{p}_j)\left(\frac{1}{m} + \frac{1}{n-m}\right)}} \xrightarrow{d} \mathcal{N}(0,1) \text{ under } H_0$$

$$= \frac{\hat{p}_j^{(m)} - \hat{p}_j^{(r)}}{\sqrt{\hat{p}_j(1-\hat{p}_j)/m}}\sqrt{\frac{n-m}{n}}$$

comparing it to a chosen critical value: z_c

Zhang, 2014 *APM*

Sequential Monitoring Procedures



the probability of incorrectly flagging an item across all of its exposures

- A series of statistical tests:
  - the number of items in CAT tests
  - the number of examinees

z_c **: case-based**

(Monte Carlo simulations)

Sequential Monitoring Procedures

- Using Response Times: method 1

– the goal is to detect a **significant decrease** in RTs

– the lognormal model:

$$f(t_{ij}|\tau_i) = \frac{1}{t_{ij}\sqrt{2\pi(1/\alpha_j)^2}} e^{-[\log t_{ij}-(\beta_j-\tau_i)]^2/[2(1/\alpha_j)^2]}$$

$$\log T_{ij}|\tau_i \sim \mathcal{N}\left[\beta_j - \tau_i, 1/\alpha_j^2\right]$$

– the average log RT of the last m examinees for item j

✓ The observation:

$$\hat{\mu}_j^{(m)} = \frac{1}{m}\sum_{i=n-m+1}^{n} \log T_{ij}$$

✓ The expectation:

$$E\left(\hat{\mu}_j^{(m)}\right) = \frac{1}{m}\sum_{i=n-m+1}^{n}(\beta_j - \tau_i), \quad Var\left(\hat{\mu}_j^{(m)}\right) = \frac{1}{m\alpha_j^2}$$

$$H_0 : \mu_j^{(m)} = \sum_{i=n-m+1}^{n}(\beta_j - \tau_i)/m$$

$$H_1 : \mu_j^{(m)} < \sum_{i=n-m+1}^{n}(\beta_j - \tau_i)/m$$

the test statistic:

$$\frac{\hat{\mu}_j^{(m)} - \sum_{i=n-m+1}^{n}(\beta_j - \tau_i)/m}{(1/\alpha_j)/\sqrt{m}} \xrightarrow{d} \mathcal{N}(0, 1) \text{ under } H_0$$

Sequential Monitoring Procedures

- Using Response Times: method 1

– the test statistic:

$$\frac{\hat{\mu}_j^{(m)} - \sum_{i=n-m+1}^{n}(\beta_j - \tau_i)/m}{(1/\alpha_j)/\sqrt{m}} \xrightarrow{d} \mathcal{N}(0,1) \text{ under } H_0$$

– to avoid having to determine specific τ_i's for each item [$\tau_i \sim N(0, 1)$]:

$$f(t_j) = \int_{-\infty}^{\infty} f(t_j|\tau_i)g(\tau_i)\mathrm{d}\tau_i = \frac{1}{t_j\sqrt{2\pi\left(1+1/\alpha_j^2\right)}}\mathrm{e}^{-[\log t_j - \beta_j]^2/\left[2\left(1+1/\alpha_j^2\right)\right]}$$

$$\log T_j \sim \mathcal{N}[\beta_j, 1+1/\alpha_j^2]$$

this convenient formulation only holds when θ_i and τ_i are independent

✓ The expectation:

$$E\left(\hat{\mu}_j^{(m)}\right) = \beta_j, \quad Var\left(\hat{\mu}_j^{(m)}\right) = \left(1+1/\alpha_j^2\right)/m$$

$$H_0 : \mu_j^{(m)} = \beta_j$$

$$H_1 : \mu_j^{(m)} < \beta_j$$

the test statistic: $$\frac{\hat{\mu}_j^{(m)} - \beta_j}{\sqrt{\left(1+1/\alpha_j^2\right)/m}} \sim \mathcal{N}(0,1) \text{ under } H_0$$

Sequential Monitoring Procedures

- Using Response Times: method 2

– moving sample **v.s.** reference sample

✓ reference sample:

$$\hat{\mu}_j^{(r)} = \frac{1}{n-m}\sum_{i=1}^{n-m} \log T_{ij}$$

✓ the variances of log RTs:

$$\hat{\sigma}_j^{2(m)} = \frac{\sum_{i=n-m+1}^{n}\left(\log T_{ij} - \hat{\mu}_j^{(m)}\right)^2}{m-1} \quad \text{and} \quad \hat{\sigma}_j^{2(r)} = \frac{\sum_{i=1}^{n-m}\left(\log T_{ij} - \hat{\mu}_j^{(r)}\right)^2}{n-m-1}$$

➡ the pooled sample variance: $\hat{\sigma}_j^2 = \dfrac{(m-1)\hat{\sigma}_j^{2(m)} + (n-m-1)\hat{\sigma}_j^{2(r)}}{n-2}$

➡ $H_0 : \mu_j^{(m)} = \mu_j^{(r)}$

$H_1 : \mu_j^{(m)} < \mu_j^{(r)}$

the test statistic: comparing it to a chosen critical value t_c (reject H_0 when $W_j < t_c$)

$$W_j = \frac{\hat{\mu}_j^{(m)} - \hat{\mu}_j^{(r)}}{\sqrt{\hat{\sigma}_j^2\left(\frac{1}{m} + \frac{1}{n-m}\right)}} = \frac{\hat{\mu}_j^{(m)} - \hat{\mu}_j^{(r)}}{\hat{\sigma}_j/\sqrt{m}}\sqrt{\frac{n-m}{m}} \sim \mathcal{T}(n-2) \ \text{ under } H_0$$

Sequential Monitoring Procedures

- Using Responses and Response Times Jointly

– Dual Univariate (DU) Procedures

✓ **either** responses **or** RTs is sufficient evidence:

$$\text{DU - 1:} \quad \text{Flag item } j \text{ if } [(Z_j > z_c) \cap (W_j < 0)] \cup [(Z_j > 0) \cap (W_j < t_c)]$$

the insignificant result is in the direction of H_1

✓ **both** responses **and** RTs are necessary:

$$\text{DU - 2:} \quad \text{Flag item } j \text{ if } (Z_j > z_c) \cap (W_j < t_c)$$

– Single Multivariate (SM) Framework [1 = response, 2 = RTs]

✓ moving sample:

$$\hat{\boldsymbol{\mu}}^{(m)} = \begin{bmatrix} \hat{\mu}_1^{(m)} \\ \hat{\mu}_2^{(m)} \end{bmatrix} \quad \text{and} \quad \hat{\boldsymbol{\Sigma}}^{(m)} = \begin{bmatrix} \hat{\sigma}_1^{2(m)} & \hat{\sigma}_{12}^{(m)} \\ \hat{\sigma}_{12}^{(m)} & \hat{\sigma}_2^{2(m)} \end{bmatrix}$$

$\hat{\mu}_1^{(m)} \rightarrow \hat{p}^{(m)}$

$\hat{\sigma}_1^{2(m)} \rightarrow \hat{p}^{(m)}(1 - \hat{p}^{(m)})(m/(m-1))$

✓ reference sample:

$$\hat{\boldsymbol{\mu}}^{(r)} = \begin{bmatrix} \hat{\mu}_1^{(r)} \\ \hat{\mu}_2^{(r)} \end{bmatrix} \quad \text{and} \quad \hat{\boldsymbol{\Sigma}}^{(r)} = \begin{bmatrix} \hat{\sigma}_1^{2(r)} & \hat{\sigma}_{12}^{(r)} \\ \hat{\sigma}_{12}^{(r)} & \hat{\sigma}_2^{2(r)} \end{bmatrix}$$

- Using Responses and Response Times Jointly

– single multivariate (SM) framework [1 = response, 2 = RTs]

✓ moving sample:

$$\hat{\boldsymbol{\mu}}^{(m)} = \begin{bmatrix} \hat{\mu}_1^{(m)} \\ \hat{\mu}_2^{(m)} \end{bmatrix} \quad \text{and} \quad \hat{\boldsymbol{\Sigma}}^{(m)} = \begin{bmatrix} \hat{\sigma}_1^{2(m)} & \hat{\sigma}_{12}^{(m)} \\ \hat{\sigma}_{12}^{(m)} & \hat{\sigma}_2^{2(m)} \end{bmatrix}$$

✓ reference sample:

$$\hat{\boldsymbol{\mu}}^{(r)} = \begin{bmatrix} \hat{\mu}_1^{(r)} \\ \hat{\mu}_2^{(r)} \end{bmatrix} \quad \text{and} \quad \hat{\boldsymbol{\Sigma}}^{(r)} = \begin{bmatrix} \hat{\sigma}_1^{2(r)} & \hat{\sigma}_{12}^{(r)} \\ \hat{\sigma}_{12}^{(r)} & \hat{\sigma}_2^{2(r)} \end{bmatrix}$$

⬇ the mean vectors of the joint distribution: asymptotic **bivariate normality** (multivariate CLT)

✓ the unbiased pooled covariance matrix:

$$\hat{\boldsymbol{\Sigma}} = \frac{m-1}{n-2}\hat{\boldsymbol{\Sigma}}^{(m)} + \frac{n-m-1}{n-2}\hat{\boldsymbol{\Sigma}}^{(r)}$$

✓ the two-sample Hotelling's T^2 statistic:

$$T^2 = \left[\hat{\boldsymbol{\mu}}^{(m)} - \hat{\boldsymbol{\mu}}^{(r)}\right]' \left[\hat{\boldsymbol{\Sigma}}\left(\frac{1}{m} + \frac{1}{n-m}\right)\right]^{-1} \left[\hat{\boldsymbol{\mu}}^{(m)} - \hat{\boldsymbol{\mu}}^{(r)}\right]$$

$$H_0 : \boldsymbol{\mu}^{(m)} = \boldsymbol{\mu}^{(r)}$$

$$H_1 : \mu_1^{(m)} > \mu_1^{(r)} \quad \& \quad \mu_2^{(m)} < \mu_2^{(r)}$$

$$F = \frac{n-3}{2(n-2)}T^2 \sim \mathcal{F}(2, n-3) \ \text{ under } H_0$$

comparing it to a chosen critical value F_c (reject H_0 when $F > F_c$)

Method

- Simulations based on real data (high-stakes CAT)
  - 2000 examinees
  - item pool: 500 items (3PLM & HLNM)
  - estimation for α_j, β_j, θ_i, and τ_i:

    MCMC routine that fixed a_j, b_j, and c_j

    center the distribution of τ_i at 0

    10,000 MCMC draws with a burn-in size of 5000

Method

- Simulation Design

  - item selection: the ASB

    5 strata of about 100 items each

  - test length: 30 items

    the first 5 were chosen randomly

  - maximum exposure rate = 0.2

  - response:

    Bernoulli distribution with $p = P_j(\theta_i)$

  - response time:

    $\log\mathcal{N}(\beta_j - \tau_i, 1/\alpha_j^2)$

  - two broad manifestations of item compromise:

    1. √ give any test-taker an opportunity to gain preknowledge of any leaked item
    2. one or more subsets of examinees gain preknowledge of different subsets of the item pool

  - the preknowledge distribution:

    1. responses:

       $P^*(X = x) = 0.999^x \cdot 0.001^{(1-x)} \quad \Leftrightarrow \quad X \sim \text{Bernoulli}(0.999)$

    2. response times:

       $$f^*(t_{ij}) = \frac{3.5}{t_{ij}\sqrt{2\pi}} e^{-3.5^2(\log t_{ij}+2)/2} \quad \Leftrightarrow \quad \log T \sim \mathcal{N}(-2, 1/3.5^2)$$

       (range from about 2 to 30 s with a mean of about 8.5 s)

Method

- Simulation Design

  - the probability of any examinee having preknowledge of any given compromised item (ψ):

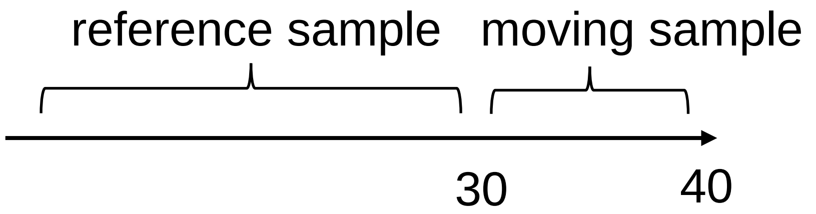
$$\psi = P(\text{preknowledge} \mid \text{compromised})$$

$$\widetilde{P}_j(\theta_i) = \psi P^*(X_{ij} = 1) + (1 - \psi) P_j(\theta_i)$$

$$\widetilde{f}_j(t_{ij}|\tau_i) = \psi f^*(t_{ij}) + (1 - \psi) f_j(t_{ij}|\tau_i)$$

  - the monitoring process:

    ✓ start for every item at the 40th exposure (e.g., m = 10)



  - compromised items:

    ✓ random quarter of the item pool (about 125 items)

    ✓ each starting at a randomized exposure count between 40 and 100

Method

- Evaluation criteria
  - C = all compromised items & F = all flagged items

1. type I error rate:

$$P(\text{Type I Error}) \approx P(F|C') = \frac{P(F \cap C')}{P(C')} = \frac{|F \cap C'|}{|C'|}$$

2. power:

$$\text{Power} \approx P(F|C) = \frac{P(F \cap C)}{P(C)} = \frac{|F \cap C|}{|C|}$$

3. the average lag:

$$\bar{L} = \frac{\sum_{j \in F \cap C}(n_j - l_j)}{|F \cap C|}$$

(change point l_j to flag point n_j)

➡ Any flagged item, whether or not in error, was recorded but otherwise **kept operational** in the item pool.

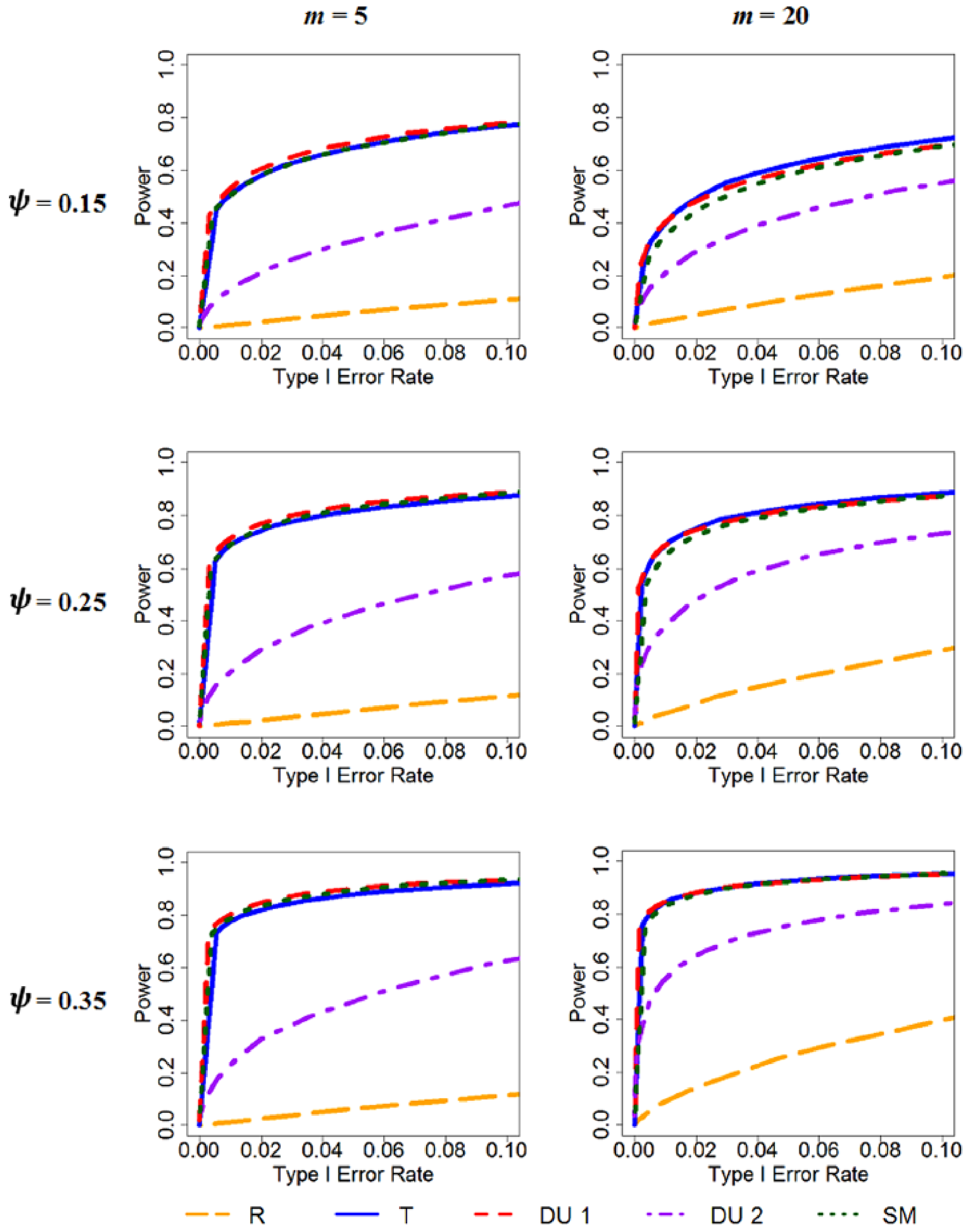
Study 1

- Purpose:
  - compared the performances of the five monitoring schemes:

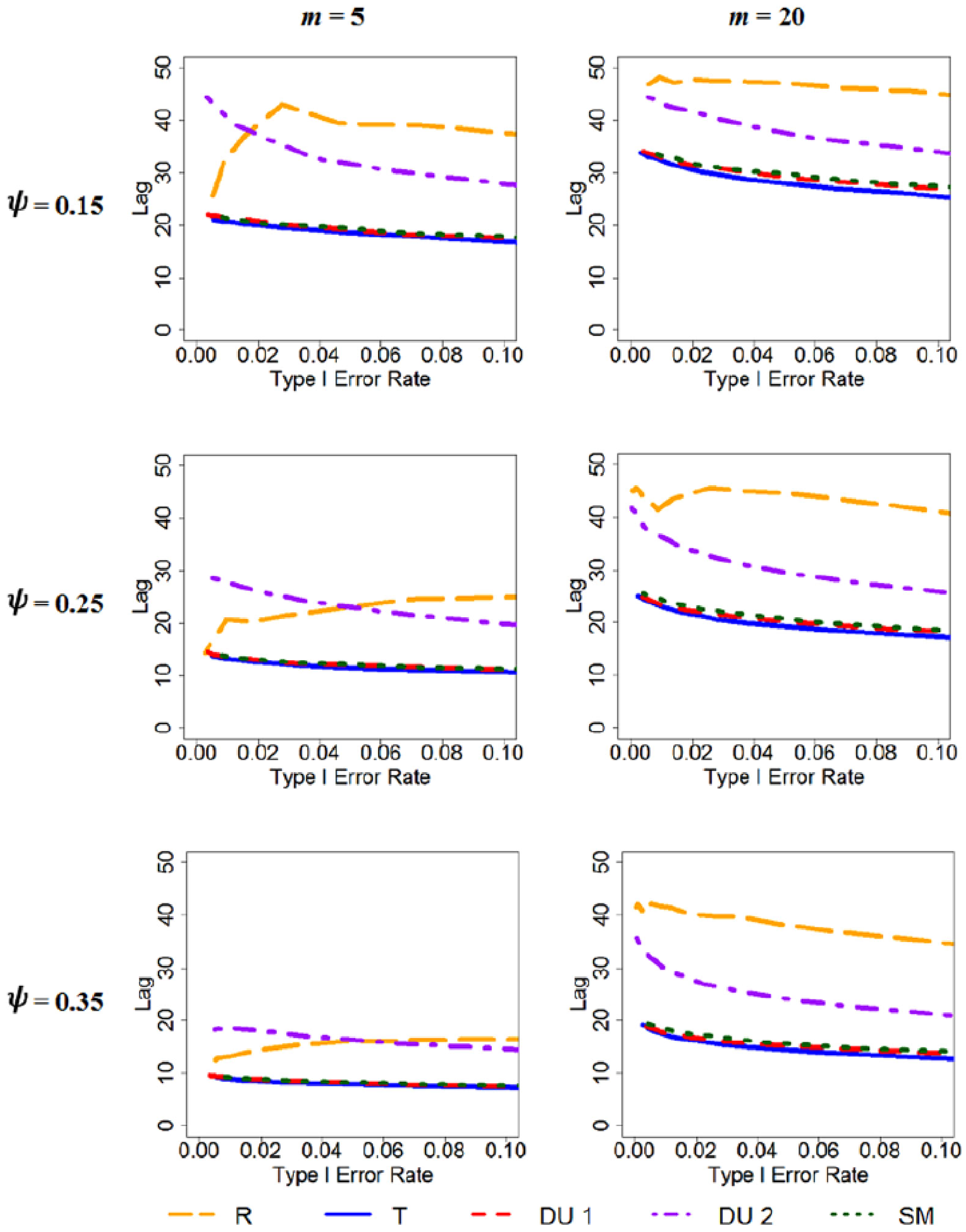
1. responses alone (R)
2. RTs alone (T)
3. dual univariate 1 (DU-1)
4. dual univariate 2 (DU-2)
5. single multivariate (SM)

- Conditions:
  - moving samples: m = 5, 20
  - preknowledge probabilities: ψ = 0.15, 0.25, 0.35
  - 100 replications

Results


m = 5
m = 20
ψ = 0.15
ψ = 0.25
ψ = 0.35
Power
Type I Error Rate
0.0 0.2 0.4 0.6 0.8 1.0
0.00 0.02 0.04 0.06 0.08 0.10
R
T
DU 1
DU 2
SM

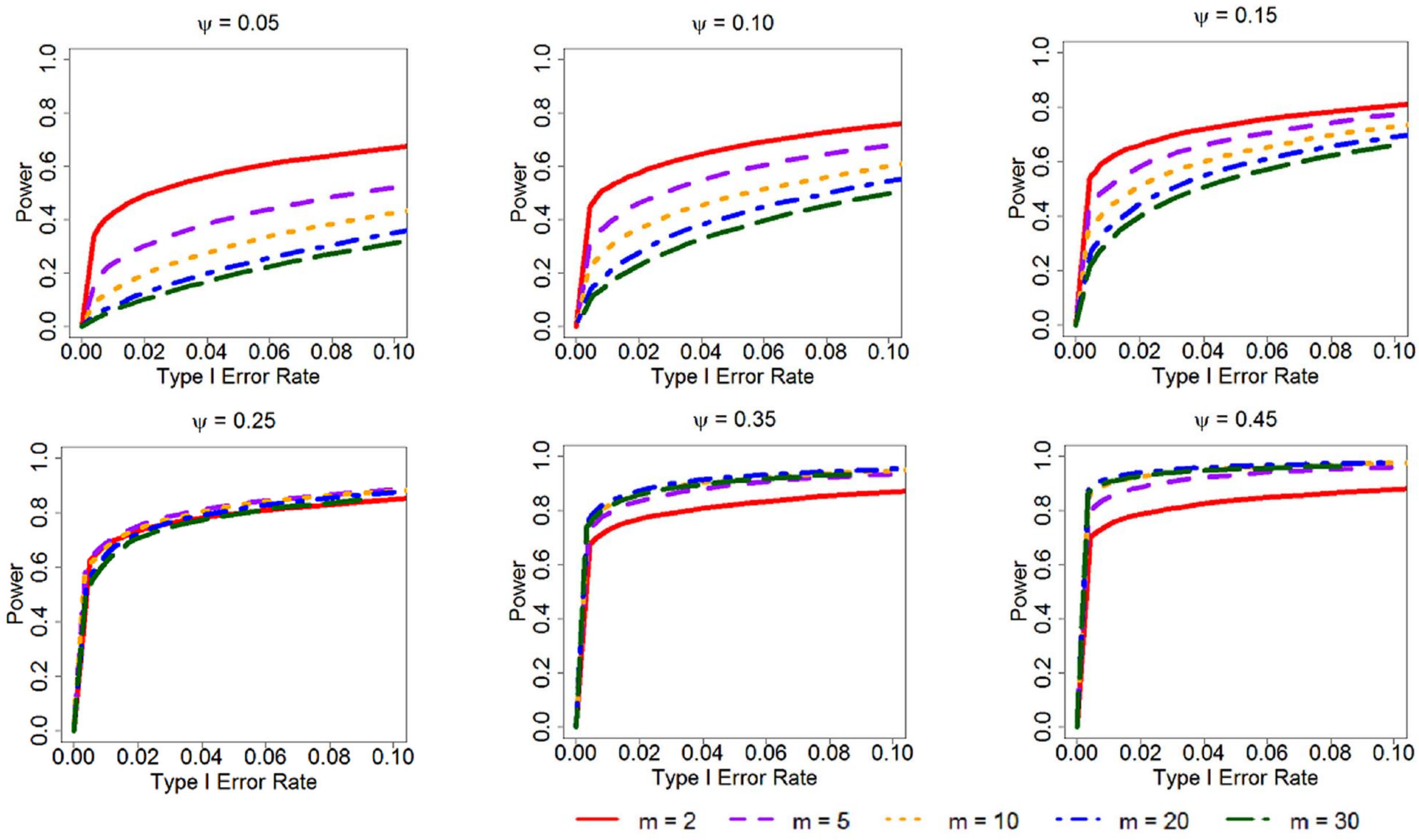
Results


m = 5
m = 20
ψ = 0.15
ψ = 0.25
ψ = 0.35
Lag
0
10
20
30
40
50
0.00
0.02
0.04
0.06
0.08
0.10
Type I Error Rate
R
T
DU 1
DU 2
SM

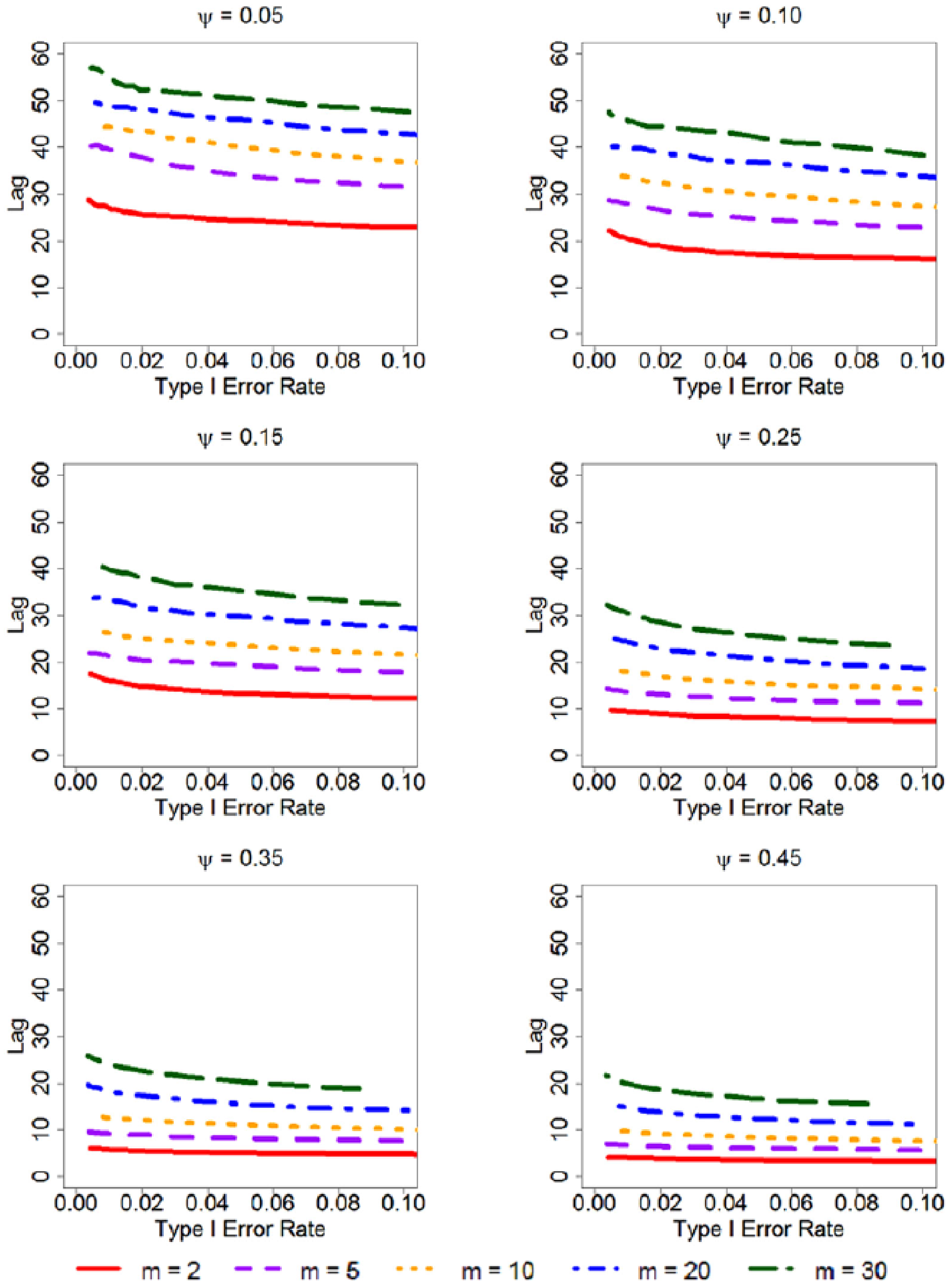
Study 2

- Purpose:
  - investigate the interaction between ψ and m
- Conditions:
  - monitoring scheme: SM
  - moving sample sizes: m = 2, 5, 10, 20, 30
  - preknowledge probabilities: ψ = 0.05, 0.10, 0.15, 0.25, 0.35, 0.45
  - 100 replications

Results


ψ = 0.05
ψ = 0.10
ψ = 0.15
ψ = 0.25
ψ = 0.35
ψ = 0.45
Power
0.0
0.2
0.4
0.6
0.8
1.0
Type I Error Rate
0.00
0.02
0.04
0.06
0.08
0.10
m = 2
m = 5
m = 10
m = 20
m = 30

Results


ψ = 0.05
ψ = 0.10
ψ = 0.15
ψ = 0.25
ψ = 0.35
ψ = 0.45
Lag
0
10
20
30
40
50
60
0.00
0.02
0.04
0.06
0.08
0.10
Type I Error Rate
m = 2
m = 5
m = 10
m = 20
m = 30

Discussion

- Both **DU-1 and SM** were shown to be equally superior over **DU-2**
- SM has two distinct advantages over DU-1:

− easier to implement

− combines all information into a single evidentiary criterion

- The choice of an appropriate moving sample size

− find the equilibrium point ψ_e:

if true ψ < ψ_e, **then** m = 2

else choose the largest m

➡ allow the moving sample size to **vary over** the course of **the monitoring process**

Discussion

- the interpretation of power:
  - a compromised item is flagged

$$\text{power} = P(F|C) \neq P(C|F)$$

= 5.5% (low base rate)

$$P(C|F) = \frac{P(F|C)P(C)}{P(F|C)P(C) + P(F|C')P(C')} = \frac{\text{Power} \times \boxed{P(C)}}{[\text{Power} \times P(C)] + [\alpha \times (1 - P(C))]}$$

$$= \frac{90\% \times 5.5\%}{(90\% \times 5.5\%) + \{5\% \times (1 - 5.5\%)\}}$$

$$\approx 50\%$$

Limitations

- The particular **lognormal distribution** used to model preknowledge RTs
- Did not consider scenarios of **drastic changes in response patterns** due to reasons unrelated to item compromise
- **the probability of item preknowledge** (ψ) was assumed to be constant & **respond** correctly with near certainty (99.9%)
- the impact of the correlation between θ and τ
- non-statistical considerations
- the classic Hotelling's T^2 statistic may not be the most appropriate choice

THANKS FOR LISTENING!

REPORTER

YINGSHI HUANG