International Association for Computerized Adaptive Testing Conference 2012
CAT – the past, present and future
12–14 August 2012, Sydney, Australia
Australian Medical Council Limited
IACAT
EPEC

Multiple Cutscore Classification Testing with the Generalized Likelihood Ratio Test

Nathan Thompson, ASC
Haniza Yon, MIMOS Berhad

Overview

- A common application of the CAT approach is to making classifications
- Often called *computerized classification testing (CCT)*
- Termination criteria are different from those used in CAT
- My research focuses on developing better termination criteria

Overview

- If using IRT, then the likelihood function is the basis (as with CAT)
- We want to determine if the LF is above or below a cutscore
- Two main ways to use the LF for that:
  - Ability confidence intervals (ACI: “adaptive mastery testing” by Kingsbury & Weiss): 1.96 SEMs above, then pass
  - Likelihood ratio (Wald, 1947): higher likelihood of pass, then pass

Overview

- The likelihood ratio approach was originally developed with a *point hypothesis* structure
- Called the sequential probability ratio test (SPRT: Reckase, 1983)
- However, a composite hypothesis structure is more conceptually relevant

Point hypothesis?

- H_0: $\theta_j = \theta_1$ (examinee is nonmaster)
- H_1: $\theta_j = \theta_2$ (examinee is master)
- θ_1 is a point (selected by you) above cutscore θ_c
- θ_2 is a point (selected by you) below cutscore θ_c
- Typically, a small constant δ is selected such that $\theta_1 = \theta_c - \delta$ and $\theta_2 = \theta_c + \delta$

Point hypothesis

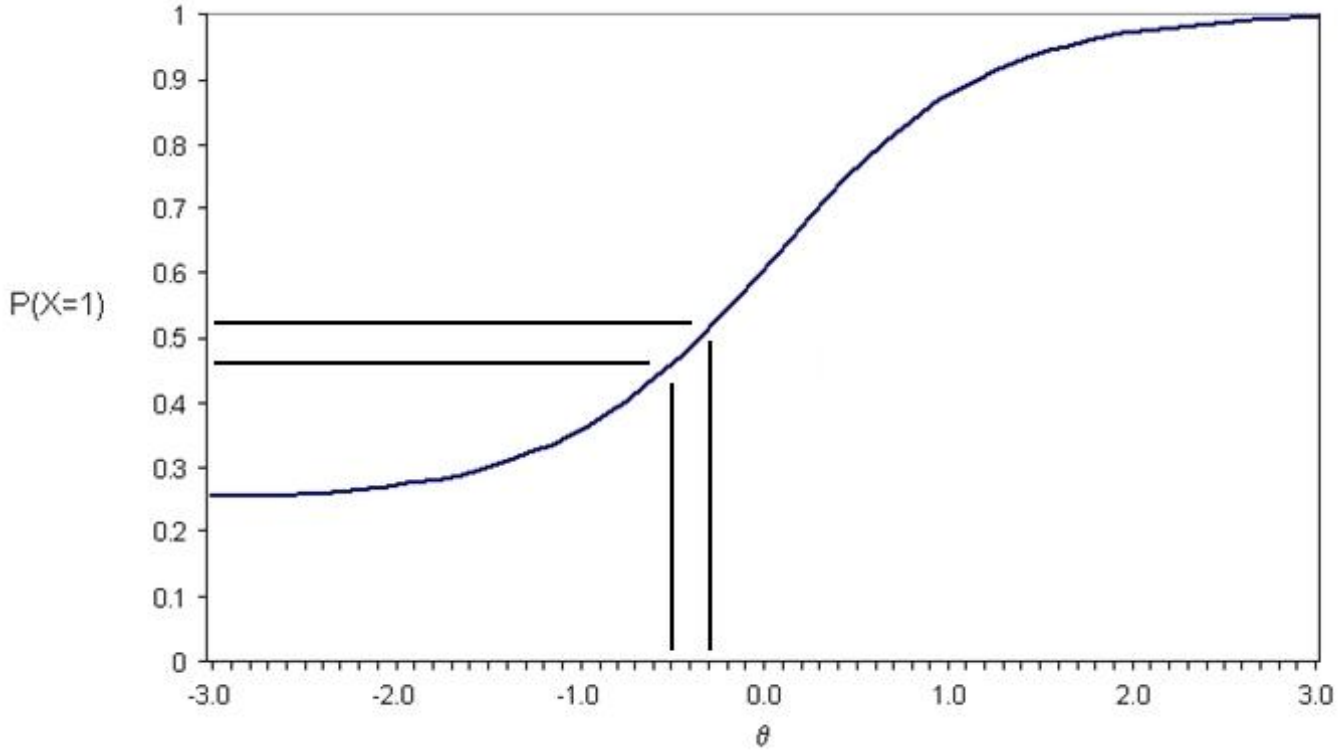
- The SPRT equation is:

$$LR_p = \frac{L(\theta = \theta_2)}{L(\theta = \theta_1)} = \frac{\prod_{i=1}^{n} P_i(X = 1 | \theta = \theta_2)^X Q_i(X = 0 | \theta = \theta_2)^{1-X}}{\prod_{i=1}^{n} P_i(X = 1 | \theta = \theta_1)^X Q_i(X = 0 | \theta = \theta_1)^{1-X}}$$

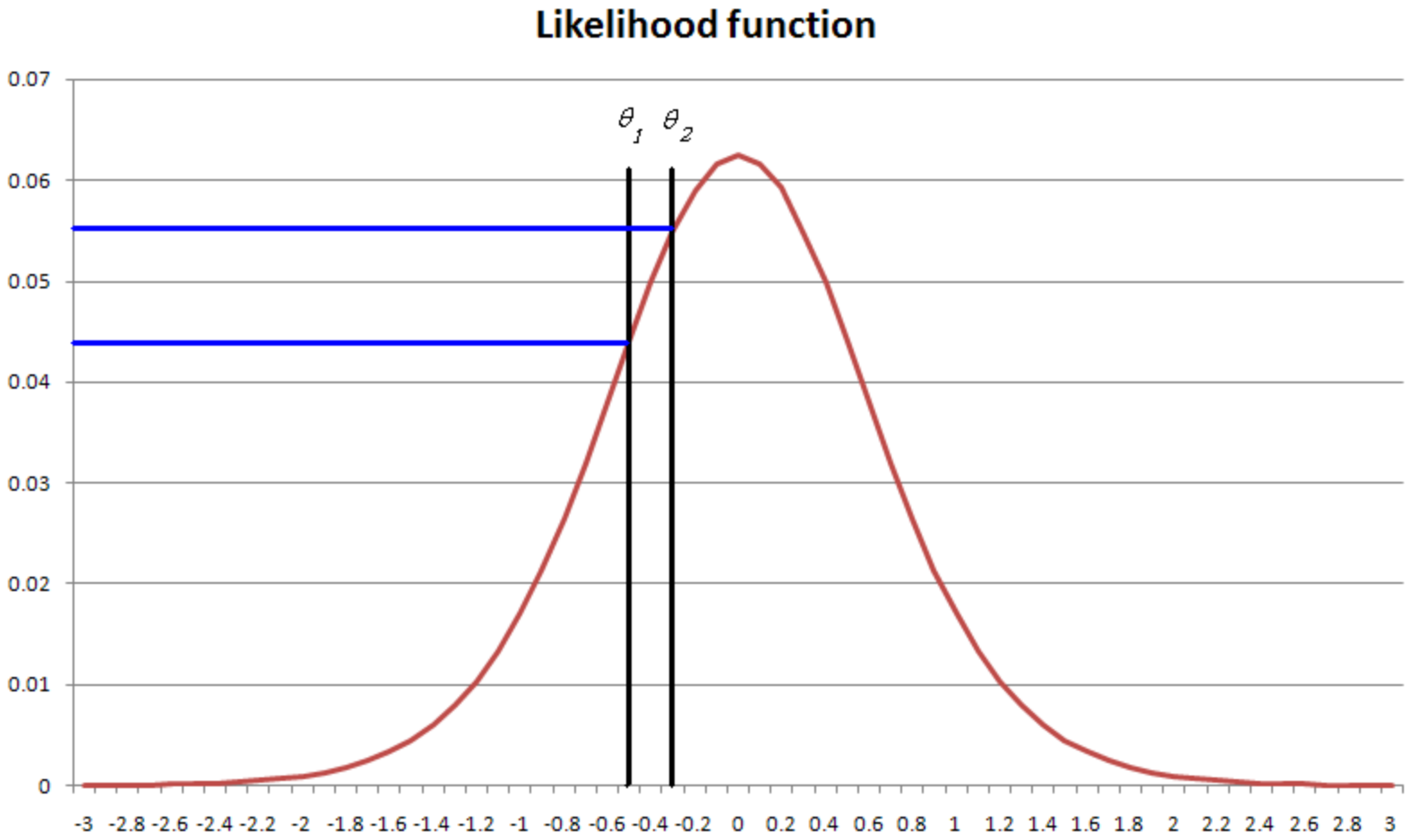
- The resulting ratio is compared to points
  - Lower decision point = $B = \beta / (1 - \alpha)$
  - Upper decision point = $A = (1 - \beta)/\alpha$
- Note that this can use CTT: originally classical difficulty statistics for distinct groups of masters and non-masters (Ferguson, 1967; also Rudner, 2002)

Point Hypothesis?

- Reckase suggested an IRT formulation, where a larger value of δ increases the difference in *P*
- Here, for one item, -0.5 & -0.3 -> 0.46 & 0.52



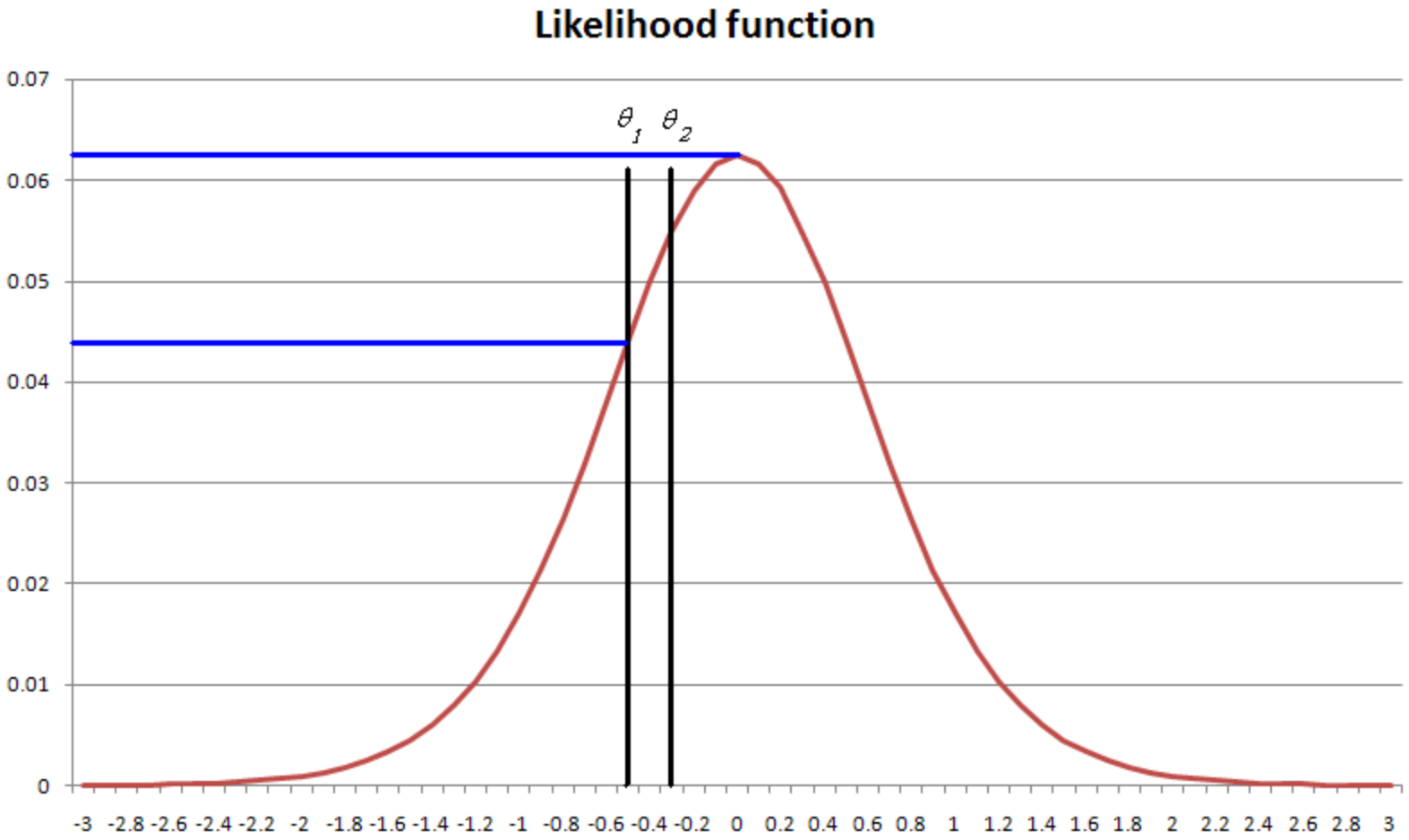
Point hypothesis: 2 items



Composite Hypothesis?

- H_0: $\theta_j \in \Theta_1$ (examinee is nonmaster)
- H_1: $\theta_j \in \Theta_2$ (examinee is master)
- Now, Θ_1 represents the *range* of θ below the cutscore, and Θ_2 represents the *range* of θ above the cutscore
- This is called the *generalized likelihood ratio test (GLRT)* – see Huang, 2004
- What is the top likelihood of each classification?

Composite hypothesis?



Thompson (2011) results

Test design	Scoring	ATL	PCC
Ability confidence intervals (ACI)	Theoretical SEM	51.65	95.73
Ability confidence intervals (ACI)	Observed SEM	54.61	95.78
Sequential probability ratio test (SPRT)	$\delta = 0.3$	39.30	95.74
Generalized likelihood ratio test (GLRT)	$\delta = 0.3$	37.62	95.73
Sequential probability ratio test (SPRT)	$\delta = 0.2$	55.77	96.21
Generalized likelihood ratio test (GLRT)	$\delta = 0.2$	48.41	96.06

- GLRT is superior
- This was only one cutscore
- What happens when two cutscores?
  - Many tests have more than one

Monte carlo simulation

- Independent variables
  - Termination criterion (ACI, SPRT, GLRT)
  - Item selection algorithm (cutscore or theta)
  - Number of cutscores (1 or 2)

MC Simulation: Dependent variables

- PCC – Percentage correctly classified (as compared to known θ value)
- ATL – Average test length, or number of items needed to make a classification
- We want PCC near nominal levels with ATL as low as possible
- Nominal accuracy: 95% for each TC
- Specify delta to produce similar PCC

MC simulation: Fixed Variables

- The bank contained 300 items.
- IRT parameters were generated
- A sample of 10,000 examinees was generated from a N(0,1) distribution.
- The cutscore was fixed at -0.5
- Min = 0, max = 200

1 cutscore results

Term	Cutoffs	Select	ATL	SDTL	PCC	Pass	Fail
ACI	1	Theta	35.88	60.53	93.19	7050	2950
ACI	1	Cutscore	48.40	64.70	95.47	6932	3068
SPRT	1	Theta	58.68	52.61	96.08	6895	3105
SPRT	1	Cutscore	56.55	54.93	95.97	6882	3118
GLRT	1	Theta	41.65	56.51	95.37	6992	3008
GLRT	1	Cutscore	41.48	54.62	95.61	6938	3062

- SPRT delta = 0.2
- GLRT delta = 0.2

1 cutscore results

Term	Cutoffs	Select	ATL	SDTL	PCC	G1	G2	G3
ACI	2	Theta	83.79	77.18	90.60	3185	3692	3123
ACI	2	Cutscore	89.68	75.26	90.91	3202	3679	3119
SPRT	2	Theta	88.94	63.81	91.93	3078	3815	3107
SPRT	2	Cutscore	88.70	62.66	91.96	3096	3820	3084
GLRT	2	Theta	56.10	62.66	90.34	3114	3647	3239
GLRT	2	Cutscore	55.55	61.16	90.74	3093	3704	3203

- SPRT delta = 0.2
- GLRT delta = 0.1

Conclusions

- GLRT and SPRT more efficient than ACI
- GLRT is more efficient than SPRT
  - Found as before for 1 cutscore
  - Difference just as strong for 2 cutscores
  - GLRT provides a distinct advantage
- Next step: try with 3 or 4 cutscores