Setting a Target Test Information Function for Assembly of IRT-Based Classification Tests

Kentaro Kato

Center for Research on Educational Testing†

Introduction

- Assembly of IRT-based tests requires a target test information function (TIF)
- Specification of the target TIF takes into consideration a desired level of estimation accuracy, overall characteristics of the item pool, empirical validation (simulation or trials and errors), etc.
- It would be useful if there is a systematic method to obtain a target TIF
- Focus on classification tests

Decision theory for pass/fail classifications

- $T(\theta)$: Target TIF
- $g(\theta) \equiv T^{-1}(\theta)$: Target "variance" function
- $\hat{\theta}$: MLE of θ, whose asymptotic sampling distribution is

$$\hat{\theta}|\theta \,\dot{\sim}\, N(\theta, g(\theta)) \tag{1}$$

- $d : \Theta \to \{\text{Pass}, \text{Fail}\}$: Decision rule

$$d(\hat{\theta}) = \begin{cases} \text{Pass}, & \hat{\theta} > \theta^* \\ \text{Fail}, & \hat{\theta} \le \theta^* \end{cases} \tag{2}$$

where θ^* is a threshold.

- $L(\theta, d)$: Loss function

	Decision	
	Fail	Pass
True State	$(\hat{\theta} > \theta^*)$	$(\hat{\theta} \le \theta^*)$
$\theta > \theta^*$	1	0
$\theta \le \theta^*$	0	1

- $R(\theta, d)$: Risk function

$$\begin{aligned} R(\theta, d) &= E_{\hat{\theta}|\theta}[L(\theta, d)] & (3) \\ &\approx \int_{\Theta} L(\theta, d)\phi(\hat{\theta}|\theta, g)\, d\hat{\theta} & (4) \\ &= \begin{cases} \Phi(\theta^*|\theta, g), & \theta > \theta^* \\ 1 - \Phi(\theta^*|\theta, g) & \theta \le \theta^* \end{cases} & (5) \end{aligned}$$

where $\phi(\cdot|\theta, g)$ and $\Phi(\cdot|\theta, g)$ are the pdf and cdf, respectively, of the normal distribution with mean θ and variance g. Here $R(\theta, d)$ is the misclassification rate given θ.

- $r(d)$: Bayes (preposterior) risk

$$\begin{aligned} r(d, h) &= E_{\theta}[R(\theta, d)] & (6) \\ &= \int_{\Theta} R(\theta, d)h(\theta)\, d\theta & (7) \end{aligned}$$

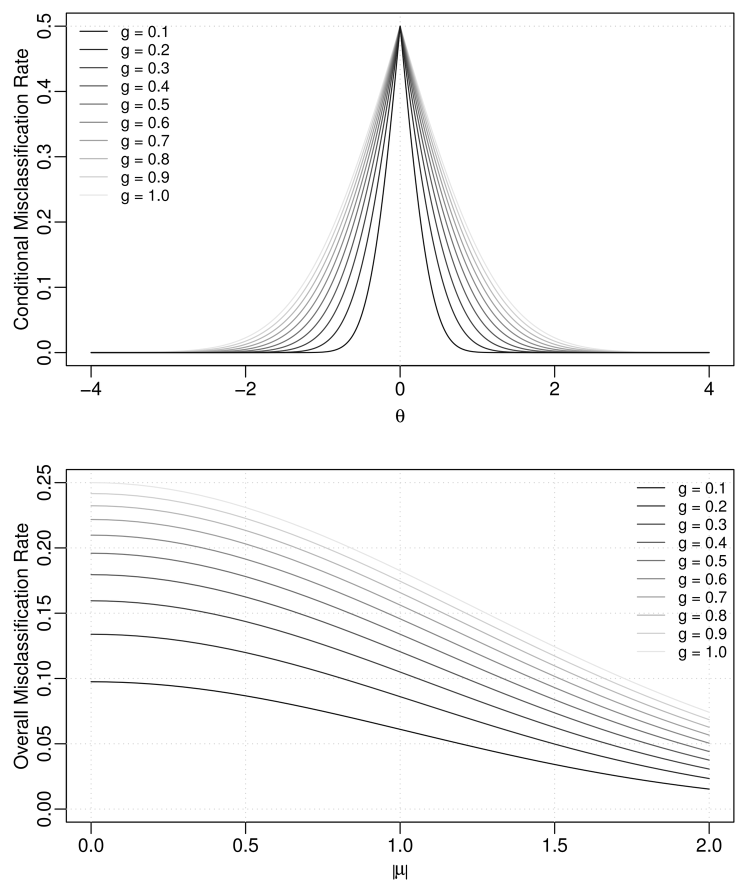
where $h(\theta)$ is a population (or prior) distribution of θ. With the above loss function, $r(d, h)$ is the overall misclassification rate.

A somewhat odd situation for a decision problem?

- In a usual decision problem, one is concerned with finding a best decision rule d (i.e., one which minimizes $r(d)$) under a given loss function L.
- In the current problem, the unknown is the sampling distribution of $\hat{\theta}$, which depends on g (or equivalently T), while d and L have already been given.
- PROBLEM
  - Specifying $g(\theta)$ which ensures the overall misclassification rate r being less than a certain value α
  - Since $g(\theta)$ affects $R(\theta, d)$ in (4) through $\phi(\hat{\theta}|\theta, g(\theta))$, the purpose is to find $g(\theta)$ which makes $r(g, h) \le \alpha$, where $r(g, h) = E_{\theta}[R(\theta, g)]$
  - For simplicity, assume $\theta^* = 0$ and $h(\theta) \stackrel{D}{=} N(\mu, 1)$

A very simplistic case

- Let $g(\theta) = \text{const.}, \forall\theta \in \Theta$ (unrealistic in IRT)
- $r(g, h)$ depends on both g and μ



Method

- It is very hard to directly deal with $r(g, h)$ (i.e., optimize $r(g, h)$ with respect to some function g). Instead, ...
  1. Set a target risk function $R(\theta)$ (i.e., conditional misclassification rate function) so that the Bayes risk r (i.e., the overall misclassification rate) with respect to a population distribution of θ falls below a prespecified level α
  2. Compute the variance function through the given values of $R(\theta)$

1. Restricting the class of risk functions

(a) It generally must be that $0 < R(\theta) \le .5, \forall\theta \in \Theta$, and especially $R(0) = .5$ (which is inevitable)

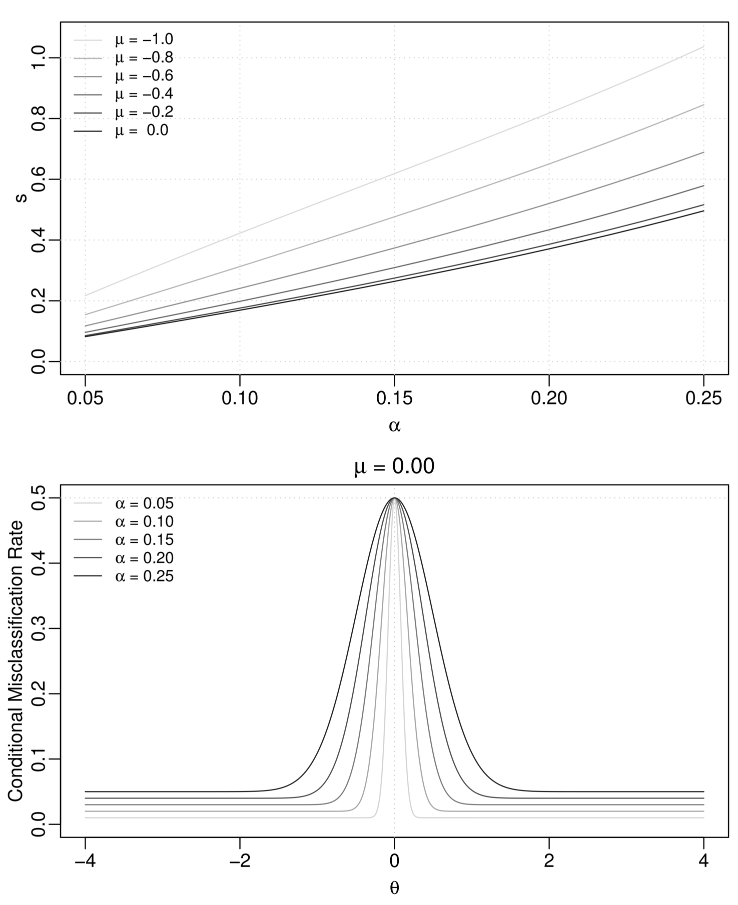
(b) A seemingly tractable and plausible form is

$$R(\theta; s, c) = \frac{1 - 2c}{2} \exp\left(-\frac{\theta^2}{2s^2}\right) + c, \tag{8}$$

where s is the "spread" and c is the "lower asymptote." The Bayes risk is then

$$r(s, h) = \frac{1 - 2c}{2} \sqrt{\frac{s^2}{1 + s^2}} \exp\left(-\frac{\mu^2}{1 + s^2}\right) + c, \tag{9}$$

which is monotone increasing with respect to $s > 0$ for all μ and $0 \le c < \alpha$. Thus, we can find the maximum s such that $r(s, h) \le \alpha$ by standard numerical methods (currently c is set to $\alpha/5$).



2. Find values of g corresponding to the target risk function

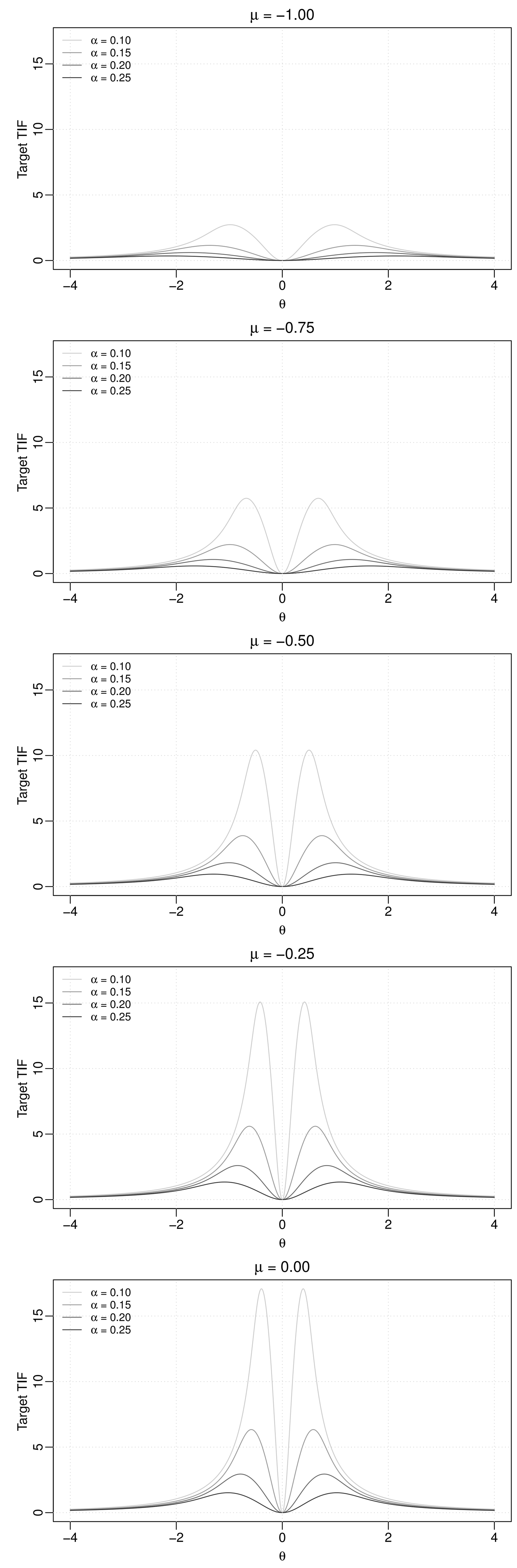
(a) For each "sampling" point θ_k, $k = 1, \ldots, q$, on Θ, the corresponding value $R_k = R(\theta_k; s, c)$ is now known. Find $z_k = \Phi^{-1}(R_k)$, where $\Phi^{-1}(\cdot)$ is the standard normal inverse CDF.

(b) Compute $g_k = (\theta_k/z_k)^2$ to obtain the "variance."

(c) $T_k = 1/g_k$, $k = 1, \ldots, q$, is the target TIF.

- Target TIFs were computed for all combinations of $\mu = -1.00, -0.75, -0.50, -0.25, 0$ and $\alpha = .10, .15, .20, .25$ by the above procedure.

Results



- Obvious tendency: more information is needed as α becomes smaller and/or μ becomes closer to the threshold
- There is a huge gap at the center of each curve
  - With the current form of risk function, high misclassification rates near the threshold are inevitable and theory tells us to "give them up"
  - Note that the misclassification rate is $.50$ at the threshold no matter how small the value of g is; actually g is discontinuous at $\theta = 0$)
- Nevertheless, the above results could serve as a good reference when setting up a target TIF
  - Besides the gaps at the center, these TIFs provide crude estimates of how large the test information should be at each θ level; this can be a good starting point when one has no idea about the necessary amount of TIF

Further considerations

- More appropriate class of risk functions
- Other decision rules, loss functions, θ estimates, ...

† Center for Research on Educational Testing, 12th Floor, Tokyo Opera City Tower, 3-20-2 Nishi-Shinjuku, Shinjuku-Ku, Tokyo, Japan 163-1412. mailto: katoken@cret.or.jp