Nonparametric comparison of regression curves for DIF detection

Adéla Hladká (neé Drabinová) & Patrícia Martinková

Institute of Computer Science of the Czech Academy of Sciences
Faculty of Mathematics and Physics, Charles University

Overview

Motivation and problem description

Estimation and test statistic

Weight functions

Simulation study

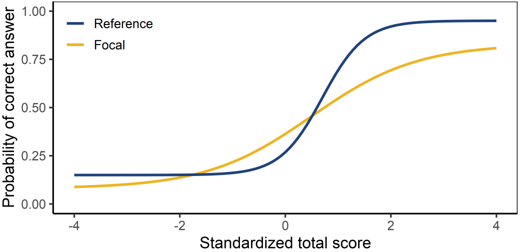
Conclusion and future work

Motivation and problem description

Motivation

DIF detection:

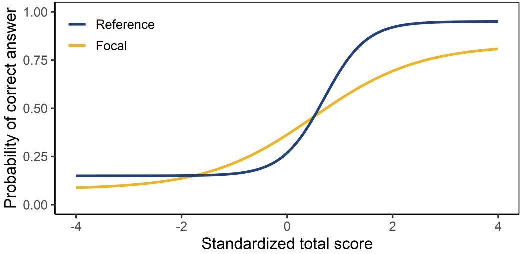
- Differences in asymptotes (guessing/inattention)



Motivation

DIF detection:

- Differences in asymptotes (guessing/inattention)



Possible issues:

- Assumption of parametric model
- Need of large sample size for 3-4PL models
- Convergence issues and instability of estimates

More general problem description

Two measurements on two populations (reference and focal)

$$\mathrm{E}(Y_R|X_R) = \mathrm{P}(Y_R = 1|X_R) = m_R(X_R),$$
$$\mathrm{E}(Y_F|X_F) = \mathrm{P}(Y_F = 1|X_F) = m_F(X_F),$$

$Y_R \in \{0, 1\}$, $Y_F \in \{0, 1\}$ (item correctness), $\mathrm{E}|Y_R| < \infty$, $\mathrm{E}|Y_F| < \infty$
X_R, X_F (standardized) total score of the test
m_R, m_F mean functions or ICCs

More general problem description

Two measurements on two populations (reference and focal)

$$\mathrm{E}(Y_R|X_R) = \mathrm{P}(Y_R = 1|X_R) = m_R(X_R),$$
$$\mathrm{E}(Y_F|X_F) = \mathrm{P}(Y_F = 1|X_F) = m_F(X_F),$$

$Y_R \in \{0,1\}$, $Y_F \in \{0,1\}$ (item correctness), $\mathrm{E}|Y_R| < \infty$, $\mathrm{E}|Y_F| < \infty$
X_R, X_F (standardized) total score of the test
m_R, m_F mean functions or ICCs

We want to test
$H_0 : m_R \equiv m_F$ vs. $H_1 : m_R \not\equiv m_F$

More general problem description

Two measurements on two populations (reference and focal)

$$\mathrm{E}(Y_R|X_R) = \mathrm{P}(Y_R = 1|X_R) = m_R(X_R),$$
$$\mathrm{E}(Y_F|X_F) = \mathrm{P}(Y_F = 1|X_F) = m_F(X_F),$$

$Y_R \in \{0,1\}$, $Y_F \in \{0,1\}$ (item correctness), $\mathrm{E}|Y_R| < \infty$, $\mathrm{E}|Y_F| < \infty$
X_R, X_F (standardized) total score of the test
m_R, m_F mean functions or ICCs

We want to test
$H_0 : m_R \equiv m_F$ vs. $H_1 : m_R \not\equiv m_F$

Observations:
$(X_{R1}, Y_{R1}), \dots, (X_{Rn_R}, Y_{Rn_R})$ (reference group)
$(X_{F1}, Y_{F1}), \dots, (X_{Fn_F}, Y_{Fn_F})$ (focal group)

Estimation and test statistic

Kernel smoothing estimate of ICC

Nearest-neighbor estimate

$$\hat{m}_R(x) = \sum_{p=1}^{n_R} Y_{Rp} W_{Rp}(x),$$

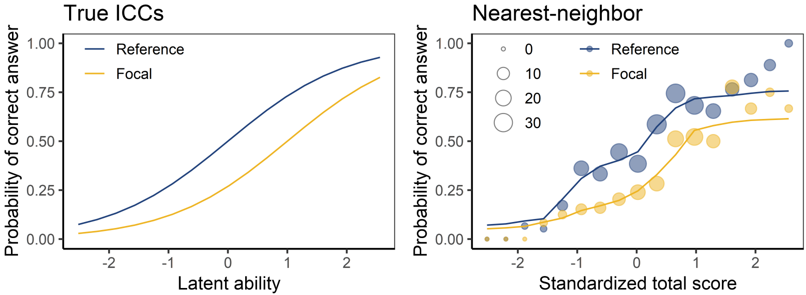
$$W_{Rp}(x) = \frac{K\left(\frac{\hat{F}_R(X_{Rp}) - \hat{F}_R(x)}{h}\right)}{\sum_{k=1}^{n_R} K\left(\frac{\hat{F}_R(X_{Rk}) - \hat{F}_R(x)}{h}\right)}$$

- K symmetric kernel function
- $\hat{F}_R(x)$ empirical distribution function of $X_{R1}, \ldots, X_{Rn_R}$
- h bandwidth

Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability & Its Applications, 9*(1), 141-142.

Srihera, R., & Stute, W. (2010). Nonparametric comparison of regression functions. *Journal of Multivariate Analysis, 101*(9), 2039–2059

Kernel smoothing estimate - example



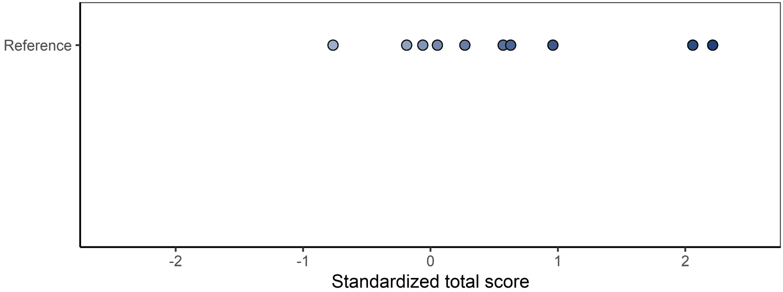
Test statistic

1. Where to compare?

Srihera, R., & Stute, W. (2010). Nonparametric comparison of regression functions. *Journal of Multivariate Analysis, 101*(9), 2039–2059

Test statistic

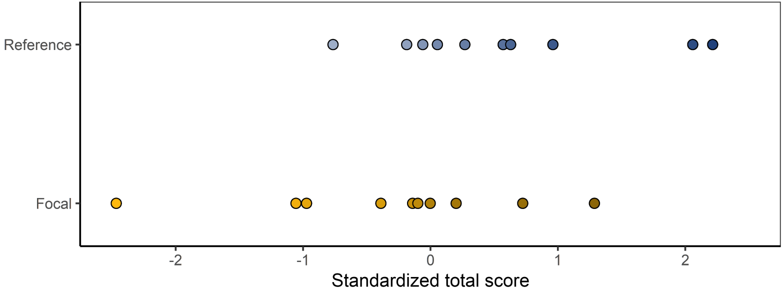
1. Where to compare?



Srihera, R., & Stute, W. (2010). Nonparametric comparison of regression functions. *Journal of Multivariate Analysis, 101*(9), 2039–2059

Test statistic

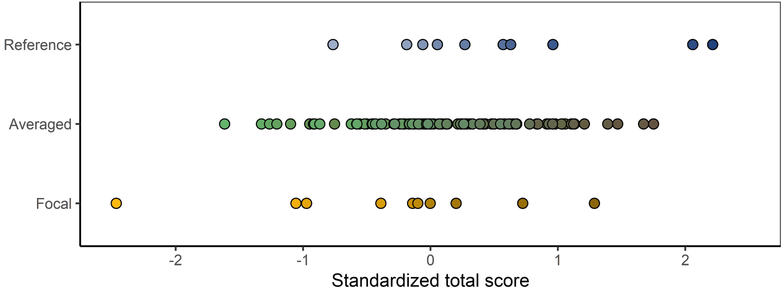
1. Where to compare?



Srihera, R., & Stute, W. (2010). Nonparametric comparison of regression functions. *Journal of Multivariate Analysis, 101*(9), 2039–2059

Test statistic

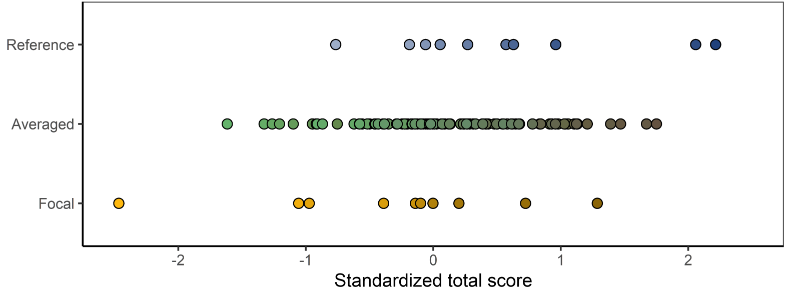
1. Where to compare?



Srihera, R., & Stute, W. (2010). Nonparametric comparison of regression functions. *Journal of Multivariate Analysis, 101*(9), 2039–2059

Test statistic

1. Where to compare?



$$\hat{T} = \frac{1}{n_R n_F} \sum_{i=1}^{n_R} \sum_{j=1}^{n_F} \left[\hat{m}_R \left(\frac{X_{Ri} + X_{Fj}}{2} \right) - \hat{m}_F \left(\frac{X_{Ri} + X_{Fj}}{2} \right) \right]$$

Srihera, R., & Stute, W. (2010). Nonparametric comparison of regression functions. *Journal of Multivariate Analysis, 101*(9), 2039–2059

Test statistic

2. Which weight function to use?

$$\hat{T} = \frac{1}{n_R n_F} \sum_{i=1}^{n_R} \sum_{j=1}^{n_F} W\left(\frac{X_{Ri} + X_{Fj}}{2}\right) \left[\hat{m}_R\left(\frac{X_{Ri} + X_{Fj}}{2}\right) - \hat{m}_F\left(\frac{X_{Ri} + X_{Fj}}{2}\right)\right]$$

- Weight function W fixed
- Can be shown that $\hat{T}$ is normally distributed

Srihera, R., & Stute, W. (2010). Nonparametric comparison of regression functions. *Journal of Multivariate Analysis, 101*(9), 2039–2059

Weight functions

Weight function

1. Fixed weight function

$$W_1(x) = 1, \quad \forall x$$

Srihera, R., & Stute, W. (2010). Nonparametric comparison of regression functions. *Journal of Multivariate Analysis, 101*(9), 2039–2059

1. Fixed weight function

$$W_1(x) = 1, \quad \forall x$$

2. Optimal weight function
   (in the sense of maximizing power of the test)

$$W_O(x) = \frac{m_R(x) - m_F(x)}{(1-\lambda)m_R(x)(1-m_R(x))\frac{e(x)}{f_R(x)} + \lambda m_F(x)(1-m_F(x))\frac{e(x)}{f_F(x)}}$$

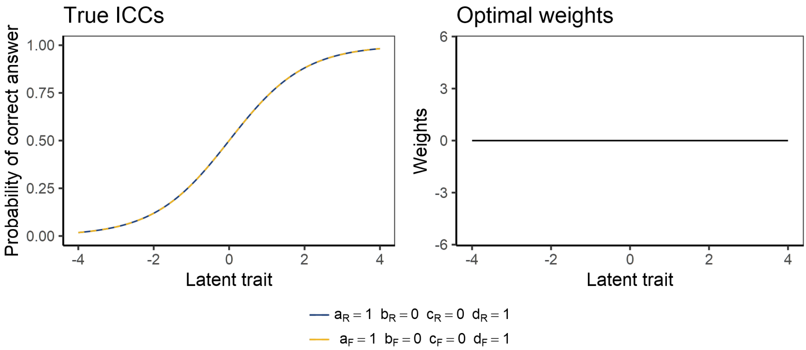
$$\lambda = \lim \frac{n_R}{n_R + n_F}$$

$f_R(x), f_F(x)$ pdf of X_R and X_F, $e(x)$ pdf of $\dfrac{X_R + X_F}{2}$

Srihera, R., & Stute, W. (2010). Nonparametric comparison of regression functions. *Journal of Multivariate Analysis, 101*(9), 2039–2059

Weight function

For 4PL IRT model with normally distributed latent trait



Weight function

3. Natural estimate of optimal weights

$$\hat{W}_O(x) = \frac{\hat{m}_R(x) - \hat{m}_F(x)}{(1-\hat{\lambda})\hat{m}_R(x)(1-\hat{m}_R(x))\frac{\hat{e}(x)}{\hat{f}_R(x)} + \hat{\lambda}\hat{m}_F(x)(1-\hat{m}_F(x))\frac{\hat{e}(x)}{\hat{f}_F(x)}}$$

- Using kernel smoothing estimates $\hat{m}_R(x)$ and $\hat{m}_F(x)$
- Test statistic is no longer normally distributed
- Asymptotic distribution not known

Wild bootstrap

Under H_0 ($m_R \equiv m_F$):

$(\hat{y}_p)_{p=1}^N$ fitted values

$(\hat{e}_p)_{p=1}^N$ residuals

Wu, C. F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *The Annals of Statistics, 14*(4), 1261-1295.

Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. *The Annals of Statistics, 21*(1), 255-285.

Wild bootstrap

Under H_0 ($m_R \equiv m_F$):

$(\hat{y}_p)_{p=1}^N$ fitted values

$(\hat{e}_p)_{p=1}^N$ residuals

Bootstrapped samples:

$$y^*_{pb} = \hat{y}_p + v_{pb}\hat{e}_p, \quad b = 1, \dots, B$$

$$v_{pb} = \begin{cases} -(\sqrt{5}-1)/2 & \text{with probability } (\sqrt{5}+1)/(2\sqrt{5}), \\ (\sqrt{5}+1)/2 & \text{with probability } (\sqrt{5}-1)/(2\sqrt{5}) \end{cases}$$

Wu, C. F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *The Annals of Statistics*, *14*(4), 1261-1295.

Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. *The Annals of Statistics*, *21*(1), 255-285.

Wild bootstrap

Under H_0 ($m_R \equiv m_F$):

$(\hat{y}_p)_{p=1}^N$ fitted values

$(\hat{e}_p)_{p=1}^N$ residuals

Bootstrapped samples:

$$y^*_{pb} = \hat{y}_p + v_{pb}\hat{e}_p, \quad b = 1, \ldots, B$$

$$v_{pb} = \begin{cases} -(\sqrt{5}-1)/2 & \text{with probability } (\sqrt{5}+1)/(2\sqrt{5}), \\ (\sqrt{5}+1)/2 & \text{with probability } (\sqrt{5}-1)/(2\sqrt{5}) \end{cases}$$

DIF detection:

- Calculation of $\hat{T}_b, b = 1, \ldots, B$
- Comparison of $\hat{T}$ using $\hat{W}_O$ and $\left(\hat{T}_b\right)_{b=1}^B$

Wu, C. F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *The Annals of Statistics, 14*(4), 1261-1295.

Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. *The Annals of Statistics, 21*(1), 255-285.

Simulation study

Design

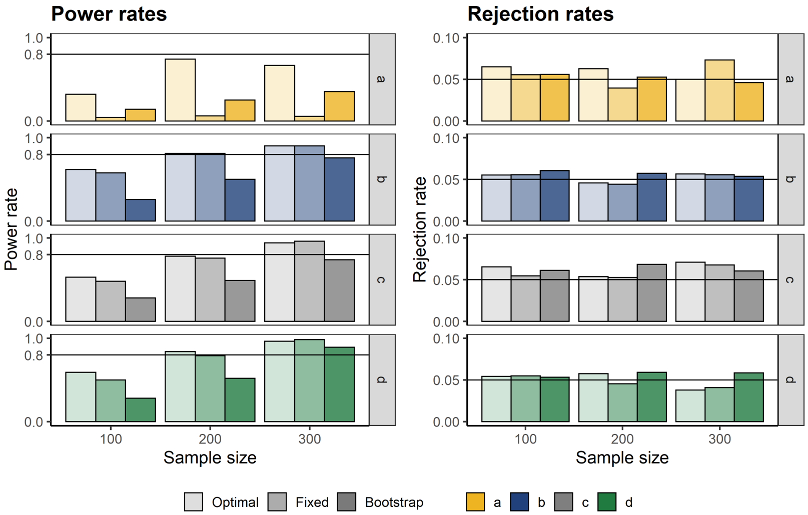
Data generation

- Data generated from true 4PL IRT model
- 20 items (1 DIF, 19 non-DIF)
- DIF caused by difference in either parameter a, b, c, or d
- Large DIF size (weighted area between ICCs $\approx$ 1.96)
- Sample sizes N = 100, 200, and 300

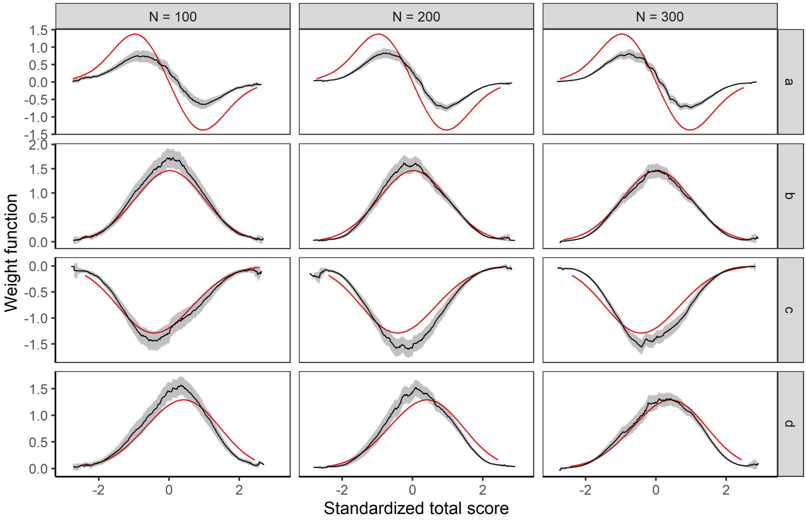
Simulation setting

- Epanechnikov kernel $K(u) = \frac{3}{4}(1 - u^2), |u| \leq 1$, $h \sim n^{-\frac{7}{24}}$
- Using optimal weights W_0, fixed weights W_1, and natural estimate $\hat{W}_0$ with bootstrap (1,000 samples)
- 100 simulation runs

Very first results

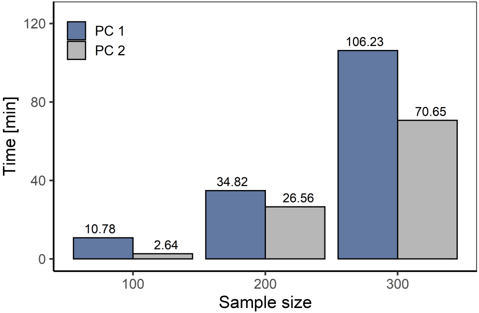


Estimates of weights



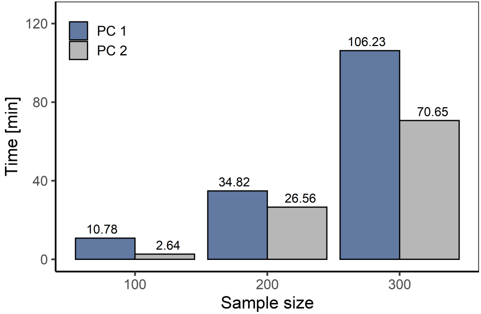
Computational complexity

Average time to test for DIF with bootstrap:



Computational complexity

Average time to test for DIF with bootstrap:



Possible reasons:

- Bootstrapping
- Length of $\left(\frac{X_{Ri}+X_{Fj}}{2}\right)_{i=1,j=1}^{nR,nF}$ vector

Conclusion and future work

Conclusion and future work

Conclusions

- Rejection rates keep at significance level
- Estimates of optimal weights $\hat{W}_O$ converge to optimal ones
- Increasing power with sample size
- DIF procedure computationally demanding

Conclusion and future work

Conclusions

- Rejection rates keep at significance level
- Estimates of optimal weights $\hat{W}_O$ converge to optimal ones
- Increasing power with sample size
- DIF procedure computationally demanding

Future work

- Complex simulation study
- Show possible superiority when true model is not 4PL IRT
- Implementation to C++ and R user-friendly functions

Questions and ideas are welcomed!

hladka@cs.cas.cz
www.cs.cas.cz/hladka/