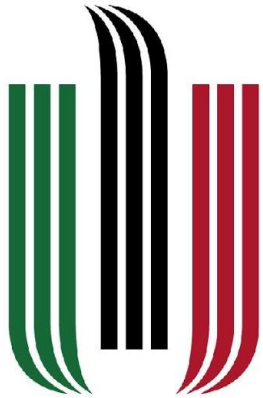


# Describing Subjective Experiment Consistency by $p$-Value P–P Plot

Jakub Nawała <jnawala@agh.edu.pl>, Lucjan Janowski, Bogdan Ćmiel, Krzysztof Rusek

ACM MM'20, October 12–16 2020, Seattle, WA, USA

Akademia Górniczo-Hutnicza im. Stanisława Staszica w Krakowie | AGH University of Science and Technology

# **Background:** General Objective

- Is the whole subjective experiment consistent?
- Is this important?
- Should I discard the whole experiment?

# **Background:** Related Work

- MOS (e.g., ITU-T P.913).
- Data consistency.[Hoßfeld2011]

**Theory**

- Scoring process modelling—Generalized Score Distribution (GSD).[Janowski2019]
- P–P plots.[Wilk1968]

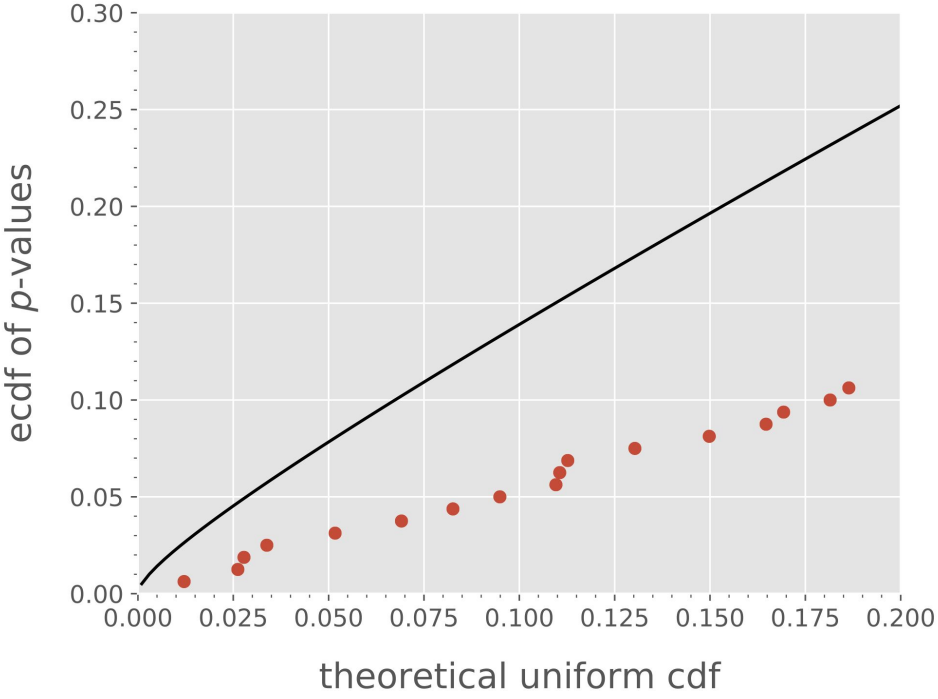
**Building Blocks**


[Hoßfeld2011] Hoßfeld, T., Schatz, R., & Egger, S. (2011). SOS: The MOS is not enough! *2011 3rd International Workshop on Quality of Multimedia Experience, QoMEX 2011, May 2014*, 131–136. https://doi.org/10.1109/QoMEX.2011.6065690

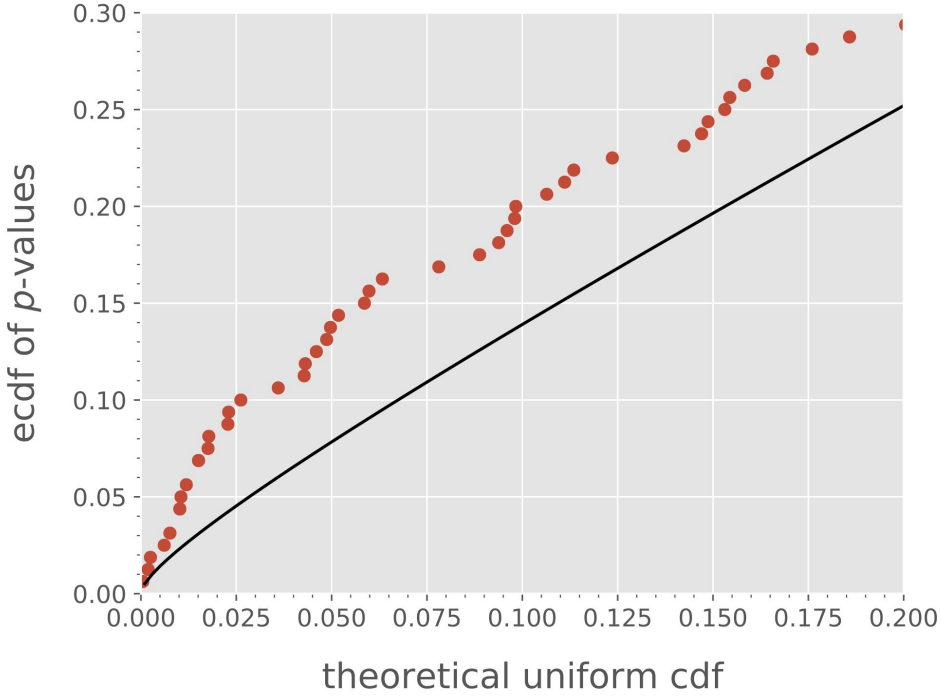
[Janowski2019] Janowski, L., Ćmiel, B., Rusek, K., Nawała, J., & Li, Z. (2019). *Generalized Score Distribution*. http://arxiv.org/abs/1909.04369

[Wilk1968] Wilk, M. B., & Gnanadesikan, R. (1968). Probability Plotting Methods for the Analysis of Data. *Biometrika*, *55*(1), 1–17.

# **Results:** The Tool




Consistent




Inconsistent

# **Results:** Three Pillars

**Theory**

Typical vs atypical score distributions.

**Detection**

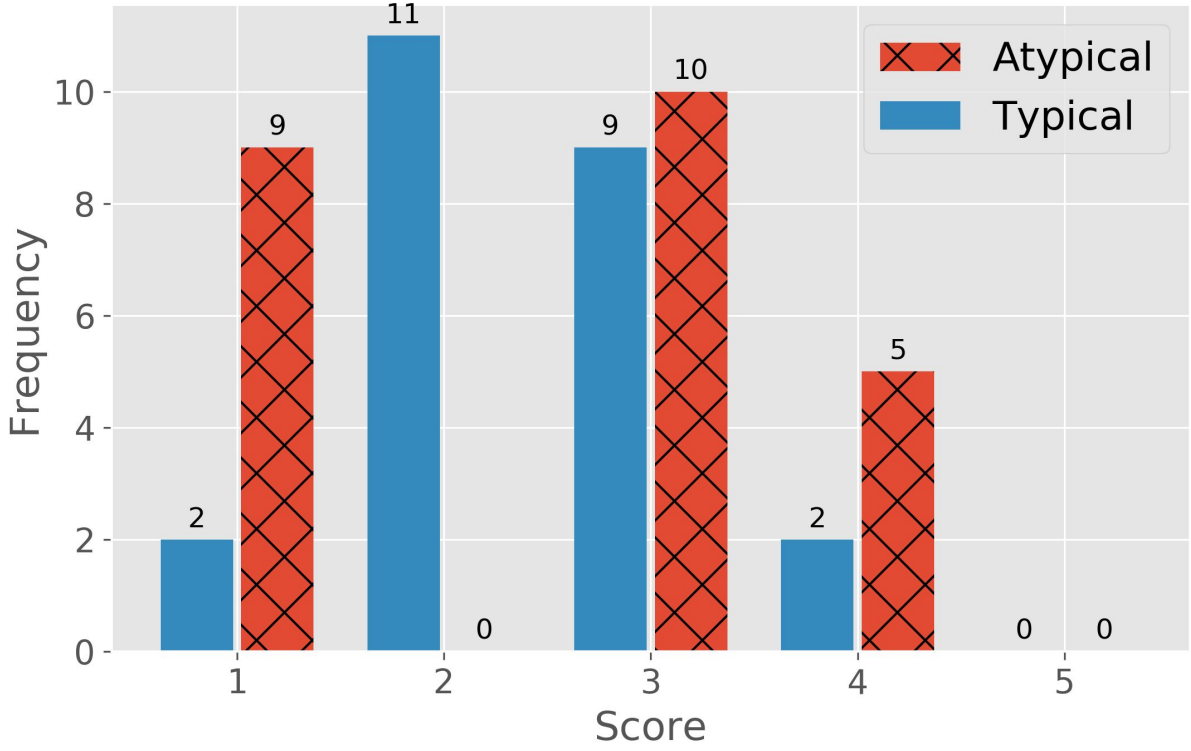
Detect atypical score distributions.

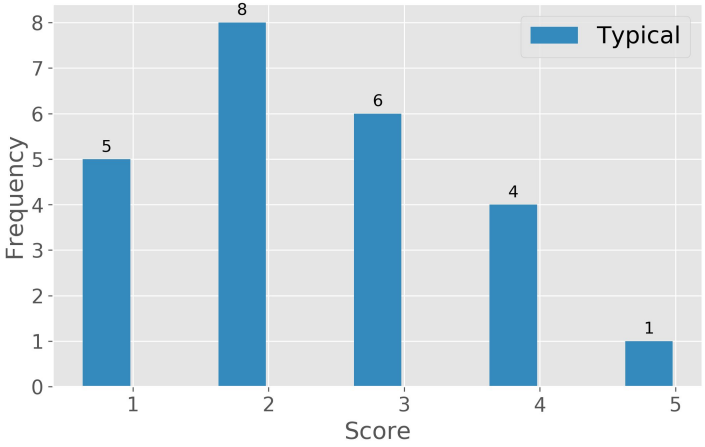
**Interpretation**

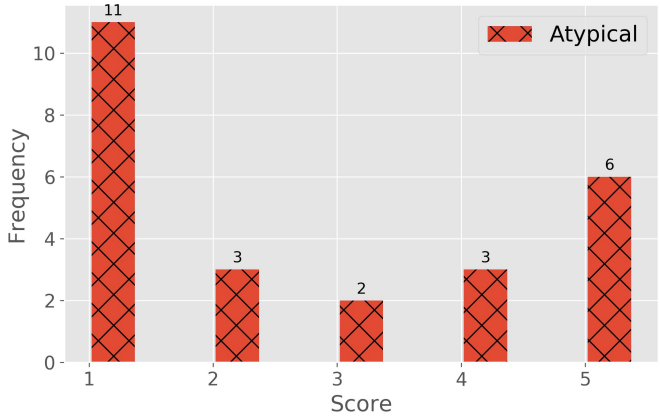
Are there too many atypical score distributions?

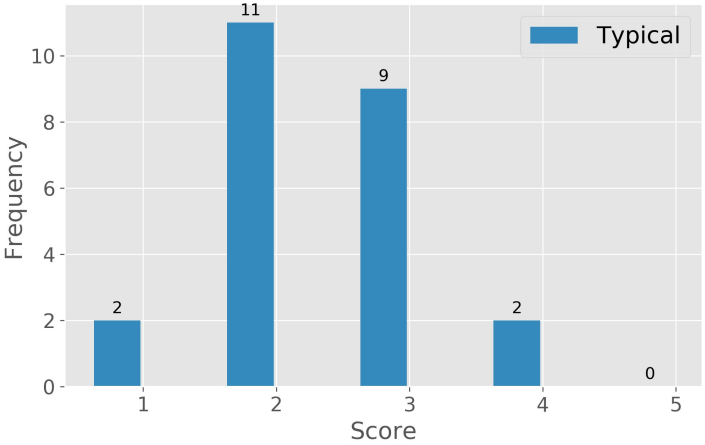
# **Results:** Theory

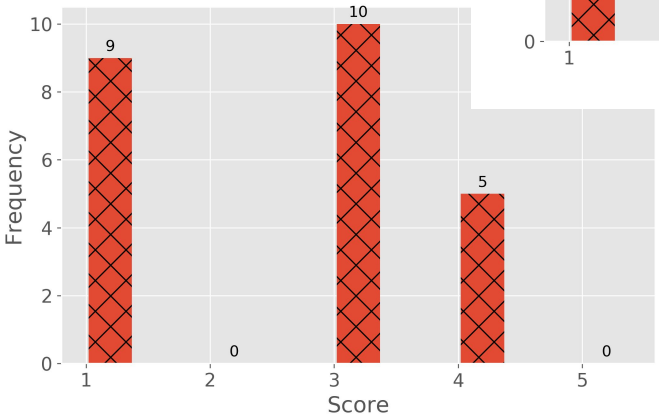
- » Quantifying QoE community's experience.
- » Score distribution > MOS.
- » Typical vs atypical.
- » GSD($\psi$, $\varrho$).

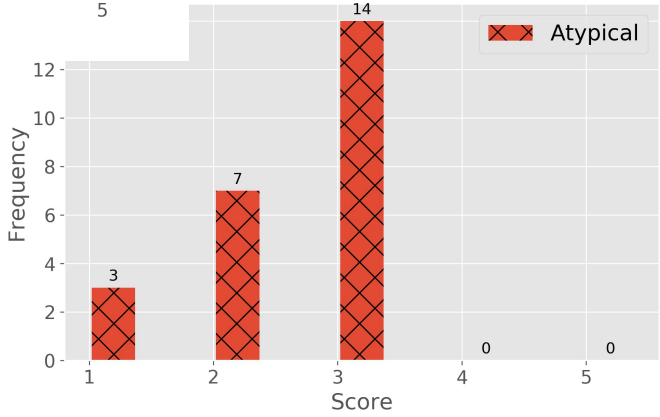


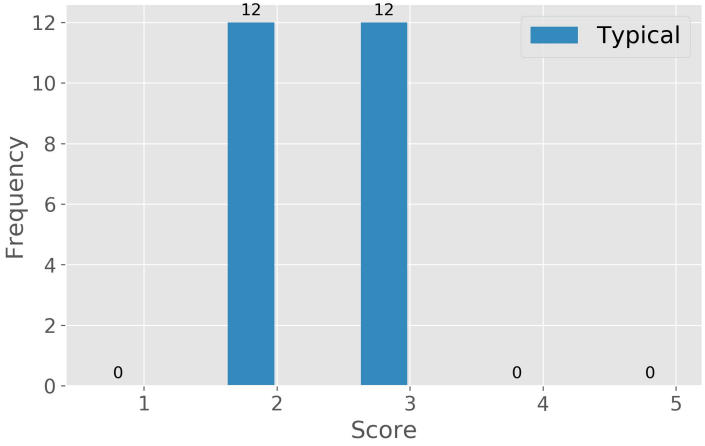

Typical
Frequency
Score
5
8
6
4
1

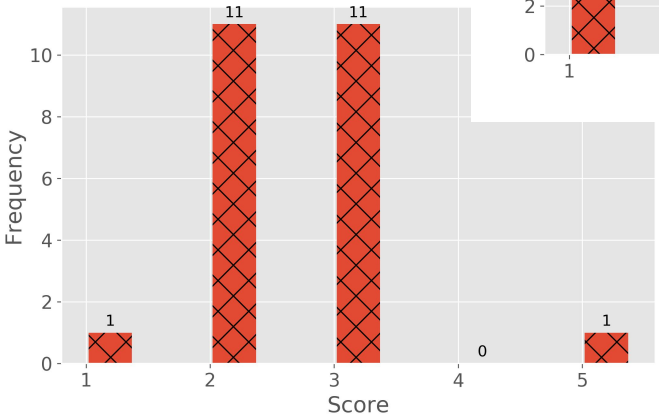

Atypical
Frequency
Score
11
3
2
3
6


Typical
Frequency
Score
2
11
9
2
0

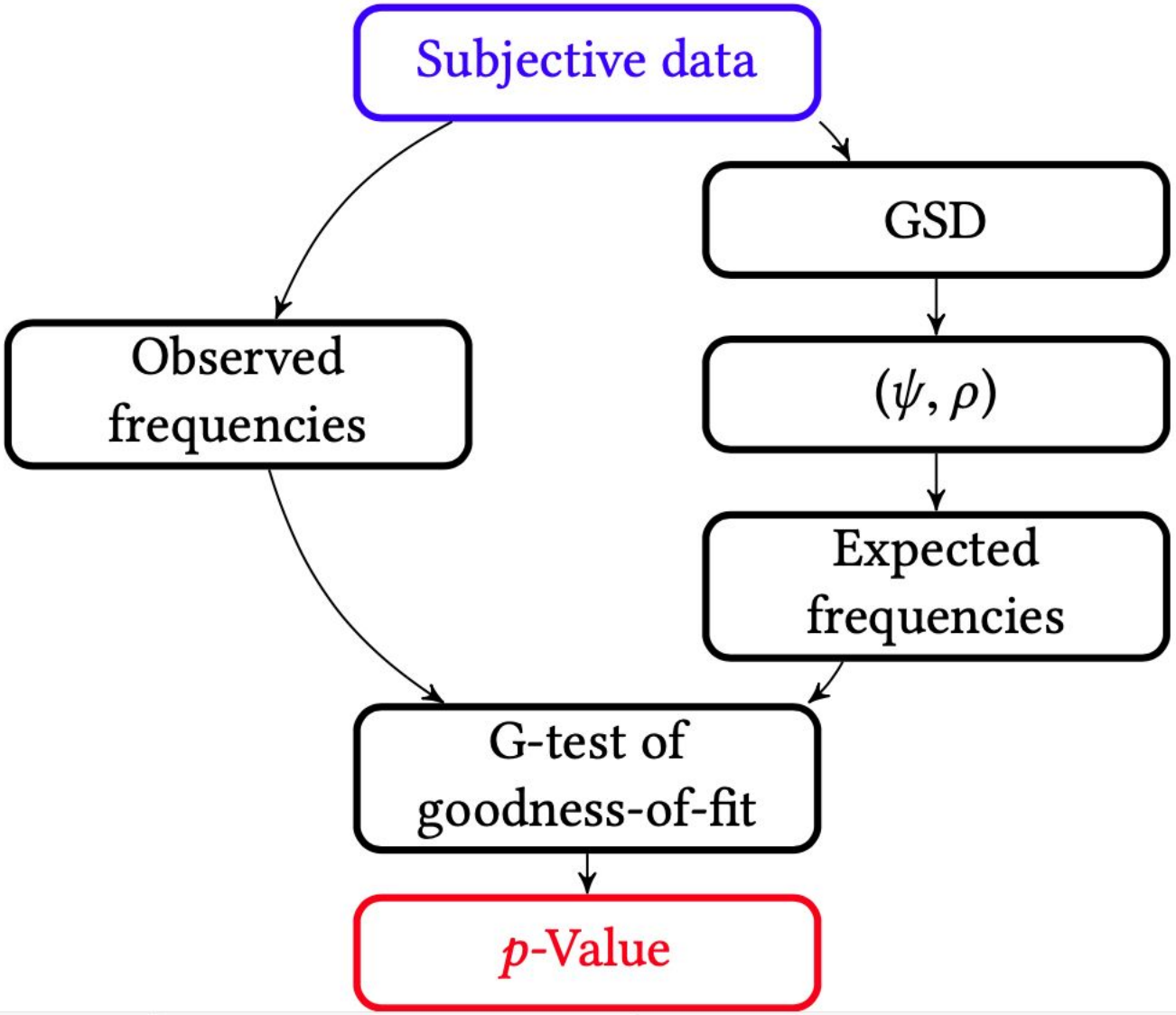

Frequency
Score
9
0
10
5
0


Atypical
Frequency
Score
3
7
14
0
0

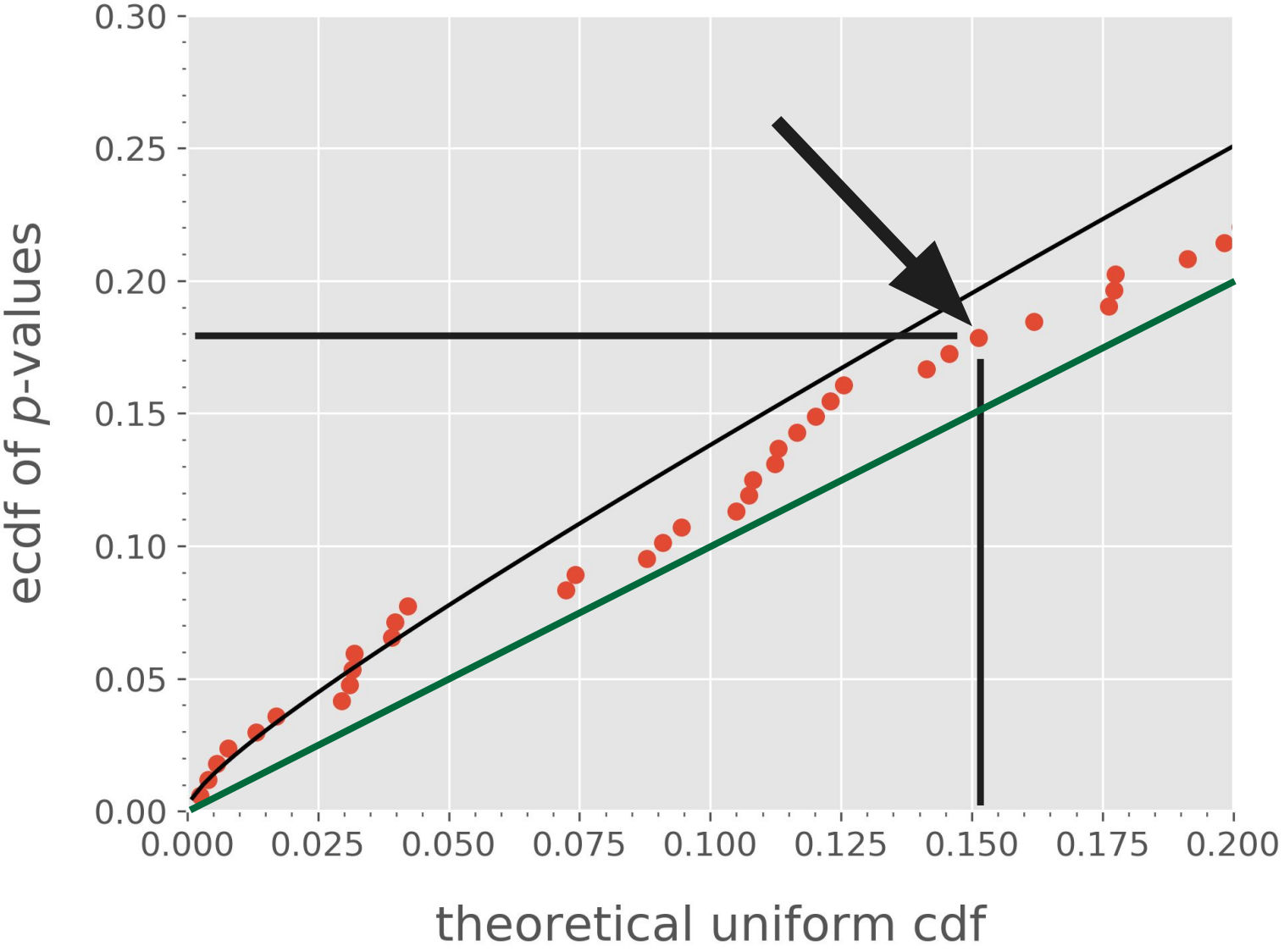

Typical
Frequency
Score
0
12
12
0
0


Frequency
Score
1
11
11
0
1

# **Results:** Detection



# **Results:** Interpretation



# Discussion & Conclusions

- A handy tool.
- Limitations.
- Future directions
  - Simulations.

# Key Takeaways

- Consistency assessment tool.
- https://github.com/Qub3k/subjective-exp-consistency-check
- Use it on your data.
- Join VQEG SAM (https://vqeg.org/).

# Thank You


- Zhi Li.
- Netlifx, Inc.
- PL-Grid Infrastructure.
- Polish Ministry of Science and Higher Education.



Jakub Nawała

jnawala@agh.edu.pl

AGH University of Science and Technology

Department of Telecommunications