Generalized Score Distribution

Bogdan Ćmiel, **Lucjan Janowski**, Jakub Nawała, Zhi Li,
Krzysztof Rusek

2019

Model for Single PVS

$$O_{ij} \sim \mathcal{N}(\psi_o, \sigma_o)$$

We have two parameters, true quality ψ_o and standard deviation for particular PVS and Subject σ_o. o stands for continuous model.

Model for Single PVS

$$O_{ij} \sim \mathcal{N}(\psi_o, \sigma_o)$$

We have two parameters, true quality ψ_o and standard deviation for particular PVS and Subject σ_o. o stands for continuous model.

After founding the new variable U has certain distribution with different parameters ψ_u and σ_u.
Note that we estimate $_u$ not $_o$ parameters! Especially we estimate ψ_u not ψ_o which was entered to the simulator or MLE function.

Model for Single PVS

$$O_{ij} \sim \mathcal{N}(\psi_o, \sigma_o)$$

We have two parameters, true quality ψ_o and standard deviation for particular PVS and Subject σ_o. o stands for continuous model.

After founding the new variable U has certain distribution with different parameters ψ_u and σ_u.
Note that we estimate $_u$ not $_o$ parameters! Especially we estimate ψ_u not ψ_o which was entered to the simulator or MLE function.
So we can validate what is the relation between ψ_o and ψ_u.

Theory

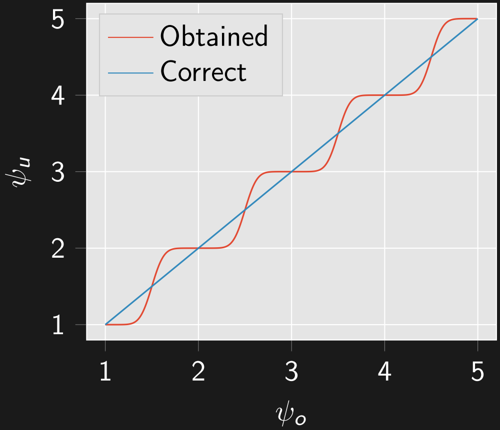
$$O_{ij} \sim \mathcal{N}(\psi_o, \sigma_o)$$

$$P(U_{ij} = k) = \begin{cases} \int_{-\infty}^{1.5} \frac{1}{\sqrt{2\pi}\sigma_o} e^{-\frac{(o-\psi_o)^2}{2\sigma_o}} & k = 1 \\ \int_{k-0.5}^{k+0.5} \frac{1}{\sqrt{2\pi}\sigma_o} e^{-\frac{(o-\psi_o)^2}{2\sigma_o}} & k \in \{2, 3, 4\} \\ \int_{4.5}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_o} e^{-\frac{(o-\psi_o)^2}{2\sigma_o}} & k = 5 \end{cases}$$

Knowing $P(U_{ij} = k)$ we can calculate ψ_u and plot function $\psi_u(\psi_o)$ it should be $y = x$.

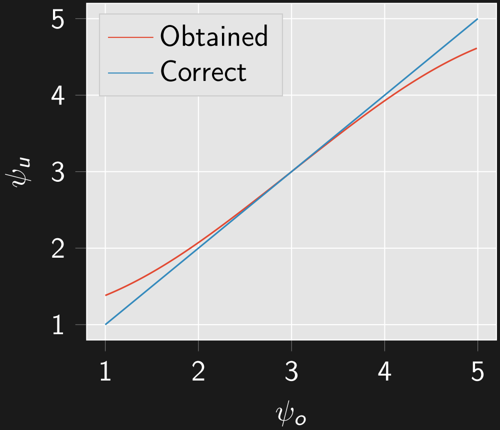
Small Standard Deviation

Let as assume that $\sigma_o = 0.1$.

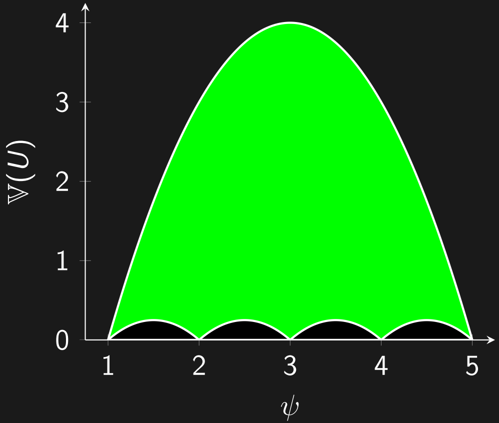


Large Standard Deviation

Let as assume that $\sigma_o = 1$.



Variance Limitation



Generalized Score Distribution

Instead of using continuous model we can use directly the discrete model.

Generalized Score Distribution

Instead of using continuous model we can use directly the discrete model.

$$P(X = 1) = F_1(\psi), P(X = 2) = F_2(\psi)$$

Example: If $\psi < 2$: $P(X = 1) = 2 - \psi$, $P(X = 2) = \psi - 1$

Instead of using continuous model we can use directly the discrete model.

$$P(X = 1) = F_1(\psi), P(X = 2) = F_2(\psi)$$

Example: If $\psi < 2$: $P(X = 1) = 2 - \psi$, $P(X = 2) = \psi - 1$

$$U \sim GSD(\psi, \rho)$$

- ψ specify the mean
- ρ specify the answers spread
  - $\rho = 1$ is the minimum variance so only $\lceil \psi \rceil$ and $\lfloor \psi \rfloor$ are possible
  - $\rho = 0$ is the maximum variance so only the minimum and the maximum value is possible

Existing Distributions

$\psi = 3.0$

0

4

Existing Distributions

$\psi = 3.0$ Φ 0 | 4

Legend:

- ○ $V_{\min}(\psi)$ - Bernoulli distribution,

Existing Distributions

Legend:

- ○ $V_{\min}(\psi)$ - Bernoulli distribution,
- ● $V_{\max}(\psi)$

Existing Distributions

$\psi = 3.0$ ⊕ ⋇ ●
0 1 4

Legend:

- $\circ$ $V_{\min}(\psi)$ - Bernoulli distribution,
- $\times$ $V_{\text{Bin}}(\psi)$ - Binomial distribution
- $\bullet$ $V_{\max}(\psi)$

Existing Distributions

Legend:

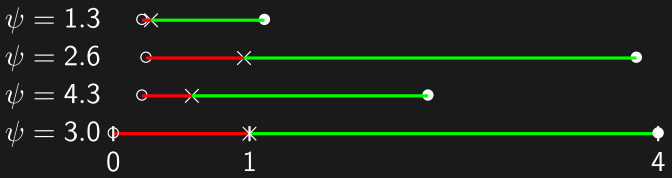
- ○ $V_{\min}(\psi)$ - Bernoulli distribution,
- × $V_{\mathrm{Bin}}(\psi)$ - Binomial distribution
- • $V_{\max}(\psi)$
- — BetaBinomial distribution

Existing Distributions

Legend:

- ○ $V_{\min}(\psi)$ - Bernoulli distribution,
- × $V_{\mathrm{Bin}}(\psi)$ - Binomial distribution
- ● $V_{\max}(\psi)$
- green line: BetaBinomial distribution
- red line: No distribution

Existing Distributions



Legend:

- ○ $V_{\min}(\psi)$ - Bernoulli distribution,
- × $V_{\text{Bin}}(\psi)$ - Binomial distribution
- ● $V_{\max}(\psi)$
- green line: BetaBinomial distribution
- red line: No distribution

Our Proposal

1. For variances larger than Binomial distribution let us use BetaBinomial Distribution
2. For variances smaller we can mix Bernoulli with Binomial. Taking more Bernoulli we increase precision, taking more of Bernoulli we have more randomness

Our Proposal

1. For variances larger than Binomial distribution let us use BetaBinomial Distribution
2. For variances smaller we can mix Bernoulli with Binomial. Taking more Bernoulli we increase precision, taking more of Bernoulli we have more randomness

$$P(X = 1) = pBer(X = 1) + (1 - p)Bin(X = 1)$$

Our Proposal

1. For variances larger than Binomial distribution let us use BetaBinomial Distribution
2. For variances smaller we can mix Bernoulli with Binomial. Taking more Bernoulli we increase precision, taking more of Bernoulli we have more randomness

$$P(X = 1) = pBer(X = 1) + (1 - p)Bin(X = 1)$$

$$P(X = 1) = p(2 - \psi) + (1 - p)\binom{4}{0}\left(\frac{\psi - 1}{4}\right)^4$$

$$H_\rho = G_\rho \; I(\rho < C(\psi)) + F_\rho \; I(\rho \geq C(\psi))$$

Equations

$$H_\rho = G_\rho \; I(\rho < C(\psi)) + F_\rho \; I(\rho \geq C(\psi))$$

Equations

$$H_\rho = G_\rho \; I(\rho < C(\psi)) + F_\rho \; I(\rho \geq C(\psi))$$

$$\begin{aligned} &P_{F_\rho}(\epsilon = k - \psi) = \\ &\quad \frac{\rho - C(\psi)}{1 - C(\psi)}[1 - |k - \psi|]_+ + \\ &\quad \frac{1 - \rho}{1 - C(\psi)} \binom{4}{k-1} \left(\frac{\psi - 1}{4}\right)^{k-1} \left(\frac{5 - \psi}{4}\right)^{5-k} \end{aligned} \tag{1}$$

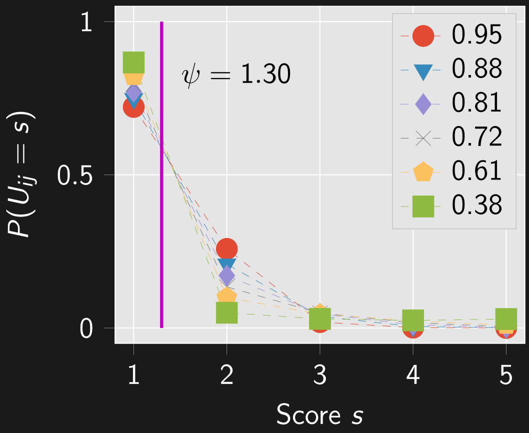
Equations

$$H_\rho = G_\rho \; I(\rho < C(\psi)) + F_\rho \; I(\rho \geq C(\psi))$$

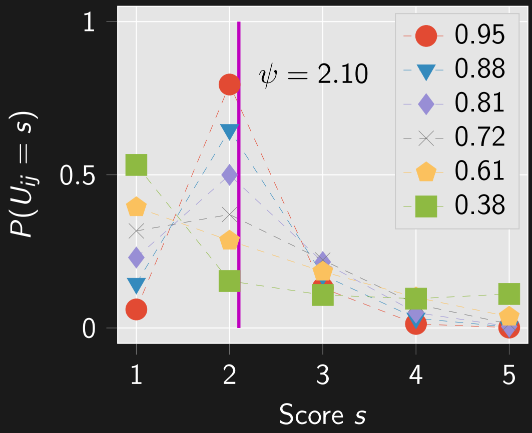
$$P_{F_\rho}(\epsilon = k - \psi) = \frac{\rho - C(\psi)}{1 - C(\psi)}[1 - |k - \psi|]_+ + \frac{1 - \rho}{1 - C(\psi)}\binom{4}{k-1}\left(\frac{\psi - 1}{4}\right)^{k-1}\left(\frac{5 - \psi}{4}\right)^{5-k} \quad (1)$$

$$P_{G_\rho}(\epsilon = k - \psi) = \binom{4}{k-1} \times \frac{\mathcal{B}\left(\frac{\rho(\psi-1)}{4(C(\psi)-\rho)} + k - 1, \frac{(5-\psi)\rho}{4(C(\psi)-\rho)} + 5 - k\right)}{\mathcal{B}\left(\frac{(\psi-1)\rho}{4(C(\psi)-\rho)}, \frac{(5-\psi)\rho}{4(C(\psi)-\rho)}\right)} \quad (2)$$

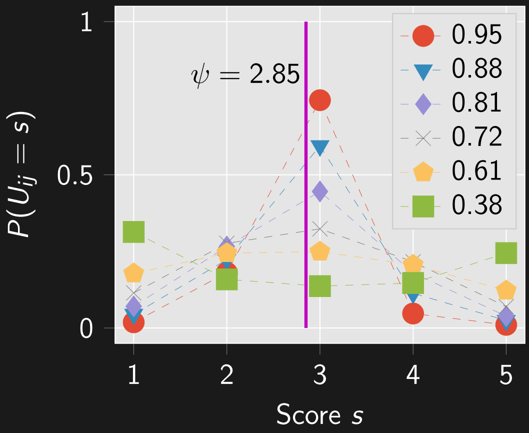
GSD Examples



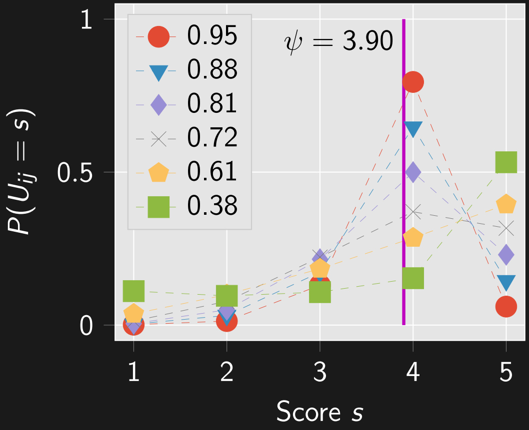
GSD Examples



GSD Examples



GSD Examples



Parameters Estimation

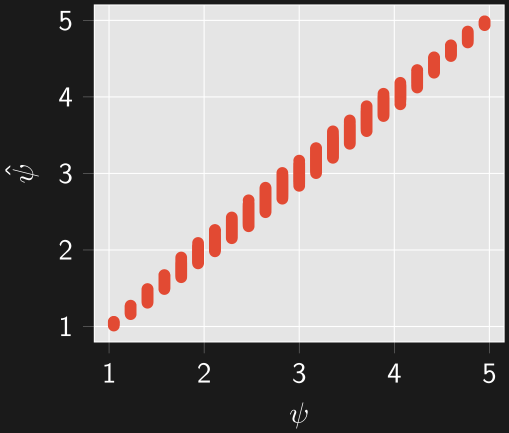
Simulation:

- a number of subjects $N = \{6, 12, 24, 48\}$,
- ψ from 1.05 to 4.95 (with 23 different values),
- ρ from 0.01 to 0.99 (with 23 different values).

Parameters Estimation

Simulation:

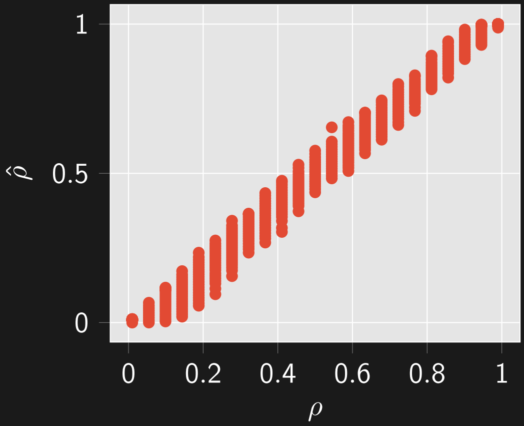
- a number of subjects $N = \{6, 12, 24, 48\}$,
- ψ from 1.05 to 4.95 (with 23 different values),
- ρ from 0.01 to 0.99 (with 23 different values).



Parameters Estimation

Simulation:

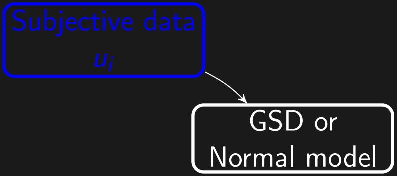
- a number of subjects $N = \{6, 12, 24, 48\}$,
- ψ from 1.05 to 4.95 (with 23 different values),
- ρ from 0.01 to 0.99 (with 23 different values).



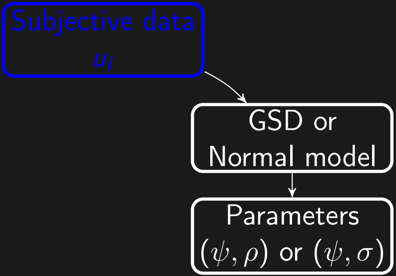
Testing Distribution

Subjective data
u_i

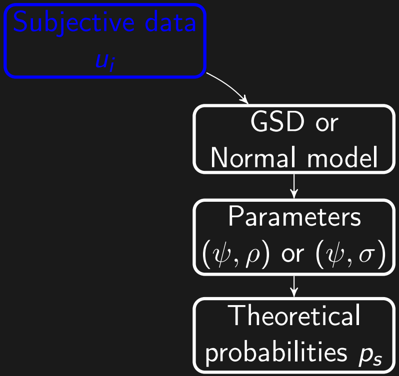
Testing Distribution



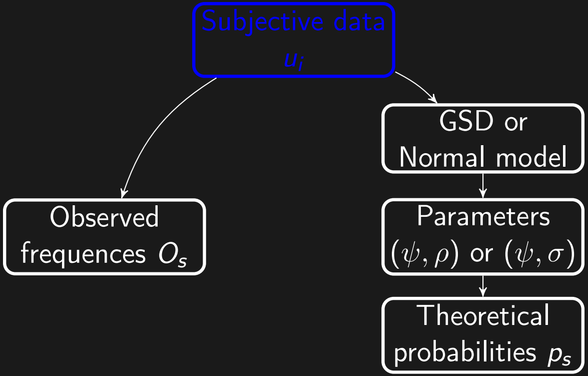
Testing Distribution



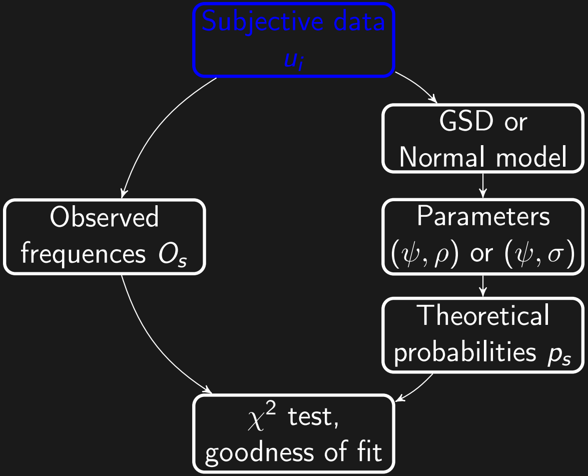
Testing Distribution



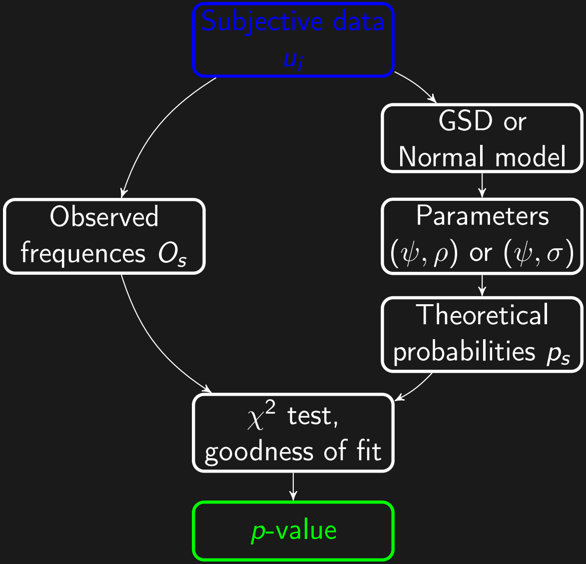
Testing Distribution



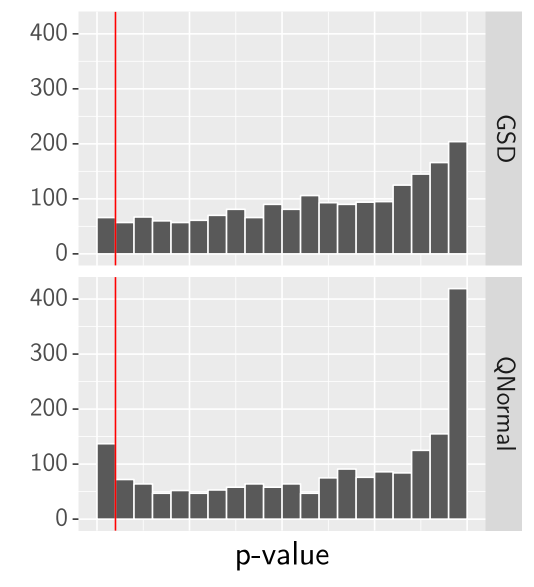
Testing Distribution



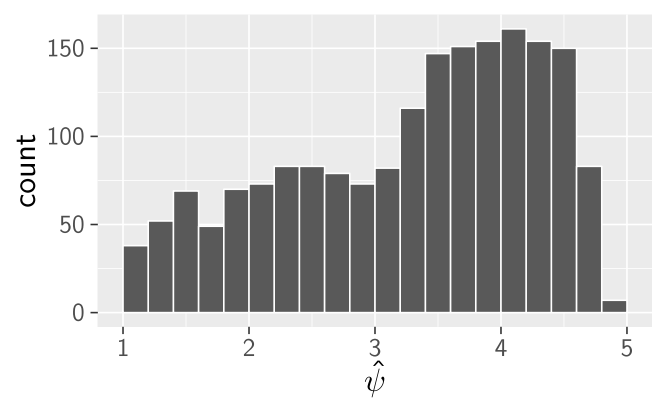
Testing Distribution



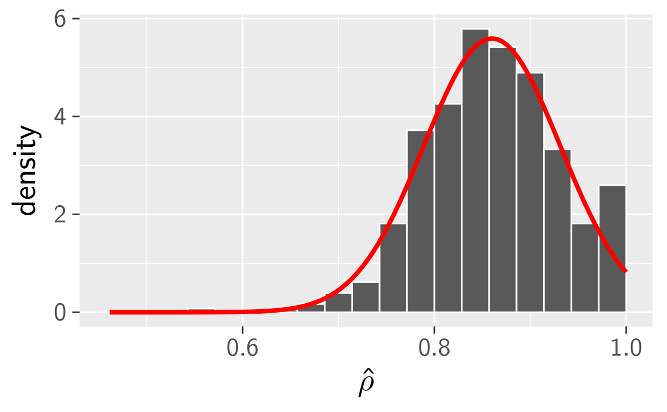
Results for 1874 Sequences



Prior Distribution ψ



Prior Distribution ρ



More Details

The full paper describing the GSD can be found here: https://arxiv.org/pdf/1909.04369.

Further Steps

- Use GSD for bootstrap
- Advance the GSD to model taking into account subject bias or error coming from subjects and work on better estimation method like Bayes
- Use GSD for different data like not pixel quality but whole movie (like imdb) or product quality
- Crate a correct test with GSD so we can test influencing factors like if gender influences the quality score
- Test which scale is correct by analyzing the answer spread

Questions