# A Thurstonian model for the Degree of Difference test with extensions to unequal variance, sequence effects and replicated data

Rune H B Christensen[1,*] John M Ennis[2] Daniel M Ennis[2]
Per B Brockhoff[1]

[1]DTU Informatics, IMM, Section for Statistics, Technical University of Denmark
[2]The Institute for Perception, Richmond, VA, USA

[*]Contact author: rhbc@imm.dtu.dk

July 11th 2012

DTU Informatics
Department of Informatics and Mathematical Modelling

# Same-Different and the Degree-of-Difference tests

2 products — 2 confusable stimuli:

- $A$ Chocolate bar (standard)
- $B$ Chocolate bar with less saturated fat

# Same-Different and the Degree-of-Difference tests

2 products — 2 confusable stimuli:

$A$ Chocolate bar (standard)

$B$ Chocolate bar with less saturated fat

Setting:

- One pair of samples evaluated at each trial
- Question: Are the samples the *same* or *different*?

# Same-Different and the Degree-of-Difference tests

2 products — 2 confusable stimuli:

- $A$ Chocolate bar (standard)
- $B$ Chocolate bar with less saturated fat

Setting:

- One pair of samples evaluated at each trial
- Question: Are the samples the *same* or *different*?

Stimuli:

- Same stimuli pairs: $AA$ and $BB$
- Different stimuli pairs: $AB$ and $BA$

# Same-Different and the Degree-of-Difference tests

2 products — 2 confusable stimuli:

$A$ Chocolate bar (standard)

$B$ Chocolate bar with less saturated fat

Setting:

- One pair of samples evaluated at each trial
- Question: Are the samples the *same* or *different*?

Stimuli:

- Same stimuli pairs: $AA$ and $BB$
- Different stimuli pairs: $AB$ and $BA$

Same-Different test:

| Same | Different |
| --- | --- |
| ☐ | ☐ |

Degree-of-Difference test:

| Same | 2 | 3 | 4 | Different |
| --- | --- | --- | --- | --- |
| ☐ | ☐ | ☐ | ☐ | ☐ |

# Characteristics of the DOD test

- An unspecified test (like Triangle, Duo-Trio, Tetrad)

# Characteristics of the DOD test

- An unspecified test (like Triangle, Duo-Trio, Tetrad)
- Only 2 samples compared at each trial

# Characteristics of the DOD test

- An unspecified test (like Triangle, Duo-Trio, Tetrad)
- Only 2 samples compared at each trial
- Easily understood test (by consumers) (O'Mahony and Rousseau, 2002)

# Characteristics of the DOD test

- An unspecified test (like Triangle, Duo-Trio, Tetrad)
- Only 2 samples compared at each trial
- Easily understood test (by consumers) (O'Mahony and Rousseau, 2002)
- No prior knowledge of products required (unlike A-not A)

# Characteristics of the DOD test

- An unspecified test (like Triangle, Duo-Trio, Tetrad)
- Only 2 samples compared at each trial
- Easily understood test (by consumers) (O'Mahony and Rousseau, 2002)
- No prior knowledge of products required (unlike A-not A)
- Response bias (like A-not A)

# Current state of research

Literature:

- Quantitative linear models by Aust et al. (1985)
- $\chi^2$ tests for replications by Bi (2002)
- ROC curve analysis by Irwin et al. (1993)

# Current state of research

Literature:

- Quantitative linear models by Aust et al. (1985)
- $\chi^2$ tests for replications by Bi (2002)
- ROC curve analysis by Irwin et al. (1993)

Gaps in our understanding:

# Current state of research

Literature:

- Quantitative linear models by Aust et al. (1985)
- $\chi^2$ tests for replications by Bi (2002)
- ROC curve analysis by Irwin et al. (1993)

Gaps in our understanding:

- Basic Thurstonian model unpublished

# Current state of research

Literature:

- Quantitative linear models by Aust et al. (1985)
- $\chi^2$ tests for replications by Bi (2002)
- ROC curve analysis by Irwin et al. (1993)

Gaps in our understanding:

- Basic Thurstonian model unpublished
- $Var(d')$ and power unknown

# Current state of research

Literature:

- Quantitative linear models by Aust et al. (1985)
- $\chi^2$ tests for replications by Bi (2002)
- ROC curve analysis by Irwin et al. (1993)

Gaps in our understanding:

- Basic Thurstonian model unpublished
- $Var(d')$ and power unknown
- Effect of $\tau$, no. categories, and ratio of $n_{same}/n_{diff}$ unknown

# Current state of research

Literature:

- Quantitative linear models by Aust et al. (1985)
- $\chi^2$ tests for replications by Bi (2002)
- ROC curve analysis by Irwin et al. (1993)

Gaps in our understanding:

- Basic Thurstonian model unpublished
- $Var(d')$ and power unknown
- Effect of $\tau$, no. categories, and ratio of $n_{same}/n_{diff}$ unknown
- No model for replications (as we have for m-AFC, Triangle, 2-AC etc.)

# Current state of research

Literature:

- Quantitative linear models by Aust et al. (1985)
- $\chi^2$ tests for replications by Bi (2002)
- ROC curve analysis by Irwin et al. (1993)

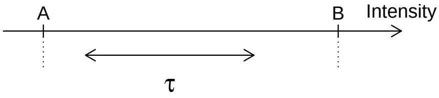
Gaps in our understanding:

- Basic Thurstonian model unpublished
- $Var(d')$ and power unknown
- Effect of $\tau$, no. categories, and ratio of $n_{same}/n_{diff}$ unknown
- No model for replications (as we have for m-AFC, Triangle, 2-AC etc.)
- No model for unequal-variance (as we have for the A-not A with sureness)

# Giving answers — $\tau$ criteria and the decision rule

# Giving answers — $\tau$ criteria and the decision rule

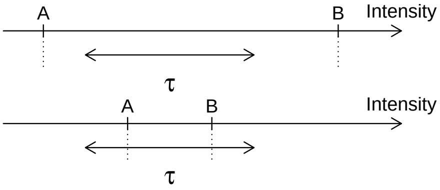
Same-Different:




$|B - A| > \tau \quad \rightarrow$ "different"

# Giving answers — $\tau$ criteria and the decision rule

Same-Different:

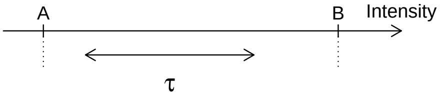



$$|B - A| > \tau \quad \rightarrow \quad \text{"different"}$$

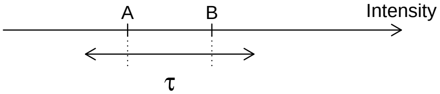
$$|B - A| < \tau \quad \rightarrow \quad \text{"same"}$$

# Giving answers — $\tau$ criteria and the decision rule

Same-Different:

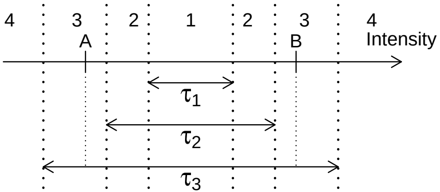



$$|B - A| > \tau \quad \rightarrow \quad \text{"different"}$$




$$|B - A| < \tau \quad \rightarrow \quad \text{"same"}$$

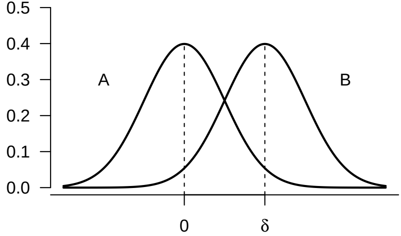
Degree of difference:



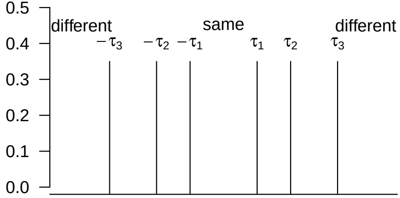

Rating scale: 1 ☐ 2 ☐ 3 ☐ 4 ☐

# Thurstonian model for the DOD test

Thurstonian distributions:

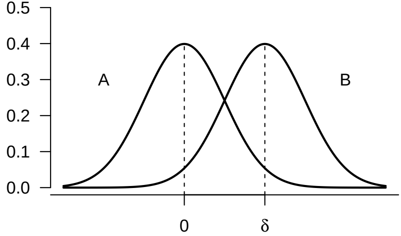



Difference distributions

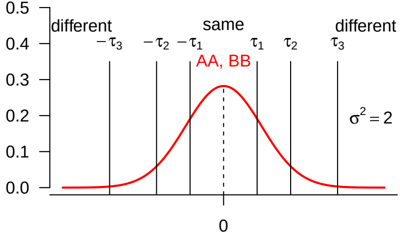


# Thurstonian model for the DOD test

Thurstonian distributions:

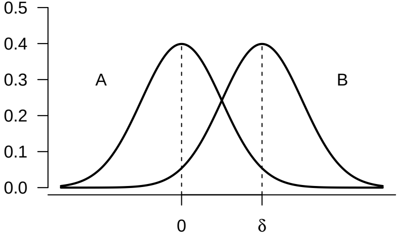



Difference distributions

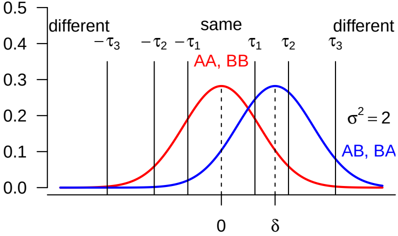


# Thurstonian model for the DOD test

Thurstonian distributions:

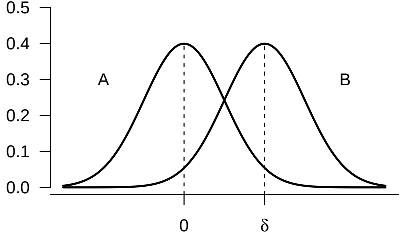



Difference distributions

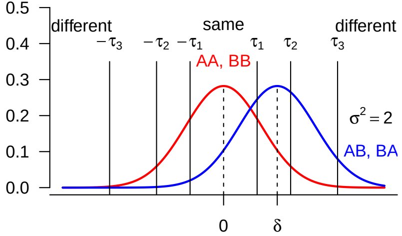


# Thurstonian model for the DOD test

Thurstonian distributions:




Difference distributions

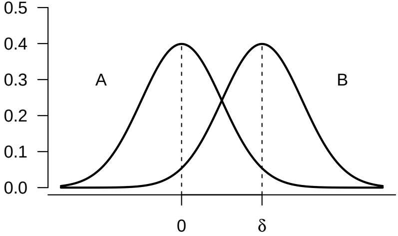



Probability of answer in the $j$th category:

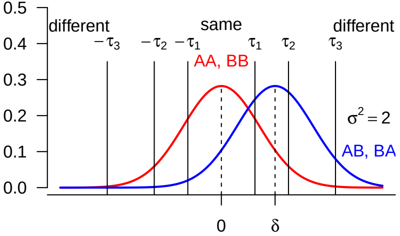
$$P(\text{"}j\text{"}|\text{Same-pair}) = f_s(\boldsymbol{\tau})$$
$$P(\text{"}j\text{"}|\text{Different-pair}) = f_d(\boldsymbol{\tau}, \delta)$$

# Thurstonian model for the DOD test

Thurstonian distributions:




Difference distributions




Probability of answer in the $j$th category:

$$P(\text{"}j\text{"}|\text{Same-pair}) = f_s(\boldsymbol{\tau})$$
$$P(\text{"}j\text{"}|\text{Different-pair}) = f_d(\boldsymbol{\tau}, \delta)$$

Maximum likelihood estimation of parameters:

$$\text{likelihood} \sim f_s(\boldsymbol{\tau}) + f_d(\boldsymbol{\tau}, \delta)$$

# Example

- 200 consumers

# Example

- 200 consumers
- Two chocolate bars (current and “healthy” alternative)

# Example

- 200 consumers
- Two chocolate bars (current and “healthy” alternative)
- 1 same-pair *or* 1 different-pair per consumer

# Example

- 200 consumers
- Two chocolate bars (current and "healthy" alternative)
- 1 same-pair *or* 1 different-pair per consumer
- 100 same-pairs + 100 different-pairs

# Example

- 200 consumers
- Two chocolate bars (current and “healthy” alternative)
- 1 same-pair *or* 1 different-pair per consumer
- 100 same-pairs + 100 different-pairs

Response scale:

| Same | | | | | | Different |
|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ |

# Example

- 200 consumers
- Two chocolate bars (current and "healthy" alternative)
- 1 same-pair *or* 1 different-pair per consumer
- 100 same-pairs + 100 different-pairs

Response scale:

| Same | | | | | | Different |
|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| □ | □ | □ | □ | □ | □ | □ |

Data obtained:

| Pair | 1 | 2 | 3 | 4 | 5 | 6 | 7 | Total |
|---|---|---|---|---|---|---|---|---|
| Same | 26 | 22 | 20 | 13 | 9 | 8 | 2 | 100 |
| Diff | 17 | 16 | 16 | 15 | 14 | 14 | 8 | 100 |

# Example: Analyses and Results

Is there a difference between products?

# Example: Analyses and Results

Is there a difference between products?

| Test | Statistic | DF | $p$-value |
|---|---|---|---|
| Pearson $\chi^2$ | 9.74 | 6 | 0.136 |

# Example: Analyses and Results

Is there a difference between products?

| Test | Statistic | DF | $p$-value |
|---|---|---|---|
| Pearson $\chi^2$ | 9.74 | 6 | 0.136 |
| $t$-test | 3.06 | 198 | 0.0025 |

# Example: Analyses and Results

Is there a difference between products?

| Test | Statistic | DF | $p$-value |
|---|---|---|---|
| Pearson $\chi^2$ | 9.74 | 6 | 0.136 |
| $t$-test | 3.06 | 198 | 0.0025 |
| Wilcoxon test | 3836.50 | – | 0.00394 |

# Example: Analyses and Results

Is there a difference between products?

| Test | Statistic | DF | $p$-value |
|---|---|---|---|
| Pearson $\chi^2$ | 9.74 | 6 | 0.136 |
| $t$-test | 3.06 | 198 | 0.0025 |
| Wilcoxon test | 3836.50 | – | 0.00394 |
| Thurstonian DOD | 3.12 | 1 | 0.00179 |

# Example: Analyses and Results

Is there a difference between products?

| Test | Statistic | DF | $p$-value |
|---|---|---|---|
| Pearson $\chi^2$ | 9.74 | 6 | 0.136 |
| $t$-test | 3.06 | 198 | 0.0025 |
| Wilcoxon test | 3836.50 | – | 0.00394 |
| Thurstonian DOD | 3.12 | 1 | 0.00179 |

Which test is the right one to use?

# Example: Analyses and Results

Is there a difference between products?

| Test | Statistic | DF | $p$-value |
|---|---|---|---|
| Pearson $\chi^2$ | 9.74 | 6 | 0.136 |
| $t$-test | 3.06 | 198 | 0.0025 |
| Wilcoxon test | 3836.50 | – | 0.00394 |
| Thurstonian DOD | 3.12 | 1 | 0.00179 |

Which test is the right one to use?

Advantages of model-based Thurstonian approach:

# Example: Analyses and Results

Is there a difference between products?

| Test | Statistic | DF | $p$-value |
| --- | --- | --- | --- |
| Pearson $\chi^2$ | 9.74 | 6 | 0.136 |
| $t$-test | 3.06 | 198 | 0.0025 |
| Wilcoxon test | 3836.50 | – | 0.00394 |
| Thurstonian DOD | 3.12 | 1 | 0.00179 |

Which test is the right one to use?

Advantages of model-based Thurstonian approach:

- Sensitive test of product differences

# Example: Analyses and Results

Is there a difference between products?

| Test | Statistic | DF | $p$-value |
|---|---|---|---|
| Pearson $\chi^2$ | 9.74 | 6 | 0.136 |
| $t$-test | 3.06 | 198 | 0.0025 |
| Wilcoxon test | 3836.50 | – | 0.00394 |
| Thurstonian DOD | 3.12 | 1 | 0.00179 |

Which test is the right one to use?

Advantages of model-based Thurstonian approach:

- Sensitive test of product differences
- Quantification of sensory intensity: $d' = 1.30(0.24)$

# Example: Analyses and Results

Is there a difference between products?

| Test | Statistic | DF | $p$-value |
|---|---|---|---|
| Pearson $\chi^2$ | 9.74 | 6 | 0.136 |
| $t$-test | 3.06 | 198 | 0.0025 |
| Wilcoxon test | 3836.50 | – | 0.00394 |
| Thurstonian DOD | 3.12 | 1 | 0.00179 |

Which test is the right one to use?

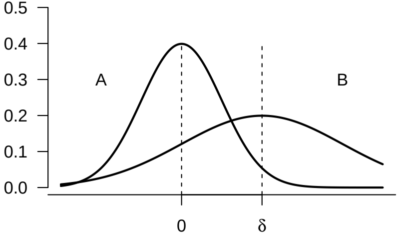
Advantages of model-based Thurstonian approach:

- Sensitive test of product differences
- Quantification of sensory intensity: $d' = 1.30(0.24)$
- Comparison of protocols

# Unequal variances and sequence effects

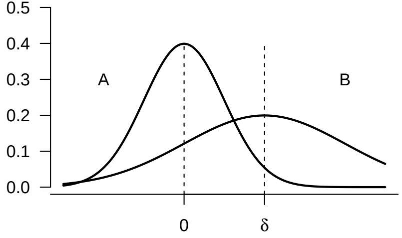
# Unequal variances and sequence effects

Unequal variances:

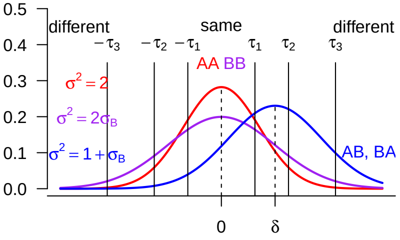


# Unequal variances and sequence effects

Unequal variances:

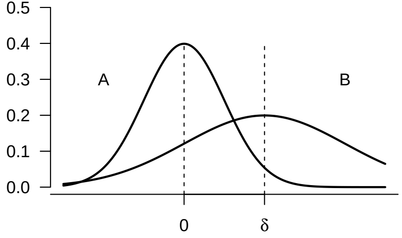



Difference distributions

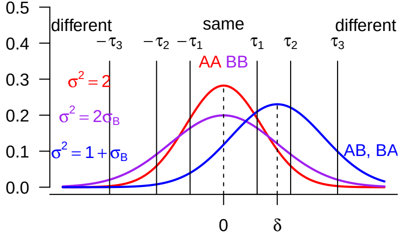


# Unequal variances and sequence effects

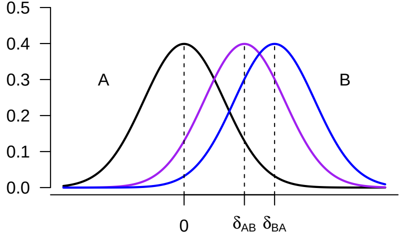
Unequal variances:




Difference distributions

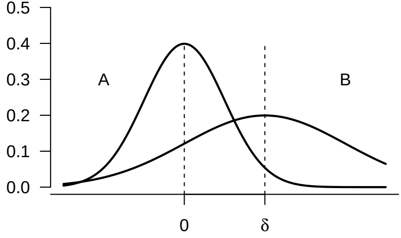



Sequence effects:

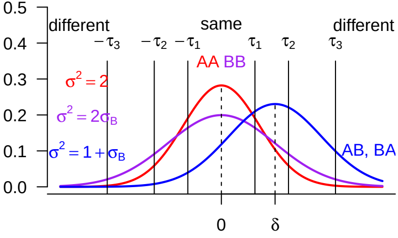


# Unequal variances and sequence effects

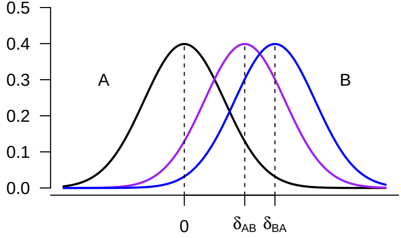
Unequal variances:



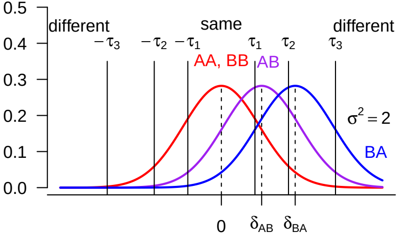

Difference distributions




Sequence effects:

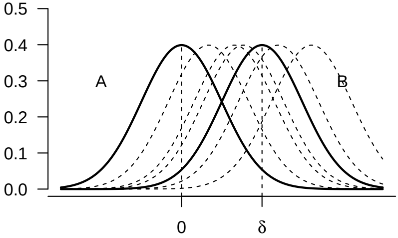



Difference distributions



# Individual differences in the DOD model

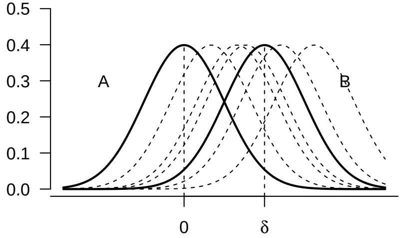
Thurstonian distributions:




- Each individual has his own $\delta$
- A model that handles replications

# Individual differences in the DOD model

Thurstonian distributions:

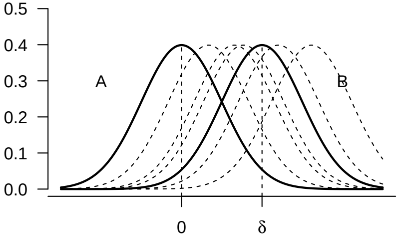



- Each individual has his own $\delta$
- A model that handles replications

Advantages of a replicated model:

# Individual differences in the DOD model

Thurstonian distributions:



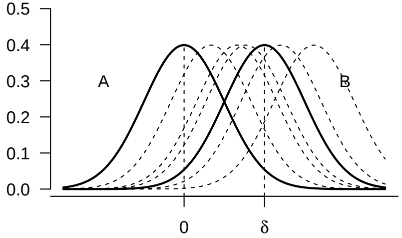

- Each individual has his own $\delta$
- A model that handles replications

Advantages of a replicated model:

- Adjust for over-dispersion

# Individual differences in the DOD model

Thurstonian distributions:



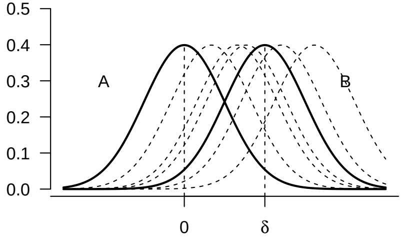

- Each individual has his own $\delta$
- A model that handles replications

Advantages of a replicated model:

- Adjust for over-dispersion
- Often more powerful tests of product differences

# Individual differences in the DOD model

Thurstonian distributions:

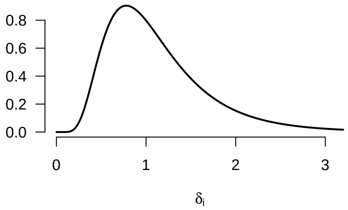



- Each individual has his own $\delta$
- A model that handles replications

Advantages of a replicated model:

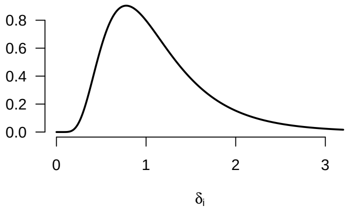
- Adjust for over-dispersion
- Often more powerful tests of product differences
- Quantification of heterogeneity

# A DOD model for replications




$$\delta_i \sim \text{log-Normal}(\delta, \sigma^2_{rep})$$

# A DOD model for replications



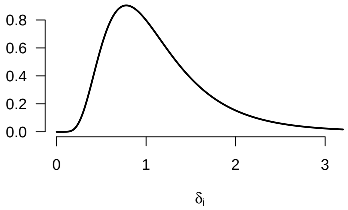

$$\delta_i \sim \text{log-Normal}(\delta, \sigma^2_{rep})$$

Probability of answer in the $j$th category:

$$P(\text{"}j\text{"}|\text{Same-pair}) = f(\boldsymbol{\tau}) \qquad \text{independent samples}$$

$$P_i(\text{"}j\text{"}|\text{Different-pair}) = f(\boldsymbol{\tau}, \delta_i) \qquad \text{NOT independent samples!}$$

# A DOD model for replications




$$\delta_i \sim \text{log-Normal}(\delta, \sigma^2_{rep})$$

Probability of answer in the $j$th category:

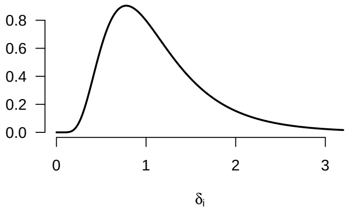
$$P(\text{"}j\text{"}|\text{Same-pair}) = f(\boldsymbol{\tau}) \qquad \text{independent samples}$$

$$P_i(\text{"}j\text{"}|\text{Different-pair}) = f(\boldsymbol{\tau}, \delta_i) \qquad \text{NOT independent samples!}$$

Computational challenge:

$$\text{log-lik} = \sum_i \log \int_0^\infty g(\boldsymbol{\tau}, \delta, \sigma_{rep}, \delta_i) \, \mathrm{d}\delta_i$$

# A DOD model for replications




$$\delta_i \sim \text{log-Normal}(\delta, \sigma^2_{rep})$$

Probability of answer in the $j$th category:

$$P(\text{"}j\text{"}|\text{Same-pair}) = f(\boldsymbol{\tau}) \qquad \text{independent samples}$$

$$P_i(\text{"}j\text{"}|\text{Different-pair}) = f(\boldsymbol{\tau}, \delta_i) \qquad \text{NOT independent samples!}$$

Computational challenge:

$$\text{log-lik} = \sum_i \log \int_0^\infty g(\boldsymbol{\tau}, \delta, \sigma_{rep}, \delta_i) \, \mathrm{d}\delta_i$$

Solution:

- Gauss-Hermite quadrature

# A replicated Degree of Difference example

25 panelists — 8 replications.

Table: Paired degree-of-difference test, data adopted from (Bi, 2002)

| | Similar | Don't know | Different | Total |
|---|---|---|---|---|
| Same pair | 45 | 40 | 15 | 100 |
| Different pair | 36 | 34 | 30 | 100 |

# A replicated Degree of Difference example

25 panelists — 8 replications.

Table: Paired degree-of-difference test, data adopted from (Bi, 2002)

| | Similar | Don't know | Different | Total |
|---|---|---|---|---|
| Same pair | 45 | 40 | 15 | 100 |
| Different pair | 36 | 34 | 30 | 100 |

Table: Tests of product differences.

| Test | $\chi^2$-value | df | $p$-value |
|---|---|---|---|
| Stuart-Maxwell test (Bi, 2002) | 3.85 | 2 | 0.149 |

# A replicated Degree of Difference example

25 panelists — 8 replications.

Table: Paired degree-of-difference test, data adopted from (Bi, 2002)

| | Similar | Don't know | Different | Total |
|---|---|---|---|---|
| Same pair | 45 | 40 | 15 | 100 |
| Different pair | 36 | 34 | 30 | 100 |

Table: Tests of product differences.

| Test | $\chi^2$-value | df | $p$-value |
|---|---|---|---|
| Stuart-Maxwell test (Bi, 2002) | 3.85 | 2 | 0.149 |
| Naive DOD test for prod | 6.10 | 1 | 0.0067 |

# A replicated Degree of Difference example

25 panelists — 8 replications.

Table: Paired degree-of-difference test, data adopted from (Bi, 2002)

| | Similar | Don't know | Different | Total |
|---|---|---|---|---|
| Same pair | 45 | 40 | 15 | 100 |
| Different pair | 36 | 34 | 30 | 100 |

Table: Tests of product differences.

| Test | $\chi^2$-value | df | $p$-value |
|---|---|---|---|
| Stuart-Maxwell test (Bi, 2002) | 3.85 | 2 | 0.149 |
| Naive DOD test for prod | 6.10 | 1 | 0.0067 |
| DOD test for reps | 5.03 | 1 | 0.0124 |

# A replicated Degree of Difference example

25 panelists — 8 replications.

Table: Paired degree-of-difference test, data adopted from (Bi, 2002)

| | Similar | Don't know | Different | Total |
|---|---|---|---|---|
| Same pair | 45 | 40 | 15 | 100 |
| Different pair | 36 | 34 | 30 | 100 |

Table: Tests of product differences.

| Test | $\chi^2$-value | df | $p$-value |
|---|---|---|---|
| Stuart-Maxwell test (Bi, 2002) | 3.85 | 2 | 0.149 |
| Naive DOD test for prod | 6.10 | 1 | 0.0067 |
| DOD test for reps | 5.03 | 1 | 0.0124 |
| DOD test prod+reps | 11.14 | 2 | 0.0038 |

# Conclusions

Main results:

# Conclusions

Main results:

- DOD protocol brought up to speed with other discrimination protocols

# Conclusions

Main results:

- DOD protocol brought up to speed with other discrimination protocols
- Thurstonian model developed

# Conclusions

Main results:

- DOD protocol brought up to speed with other discrimination protocols
- Thurstonian model developed
- Three extensions of the Thurstonian DOD-model proposed:
  - Unequal variance
  - Sequence effects
  - Replications

# Conclusions

Main results:

- DOD protocol brought up to speed with other discrimination protocols
- Thurstonian model developed
- Three extensions of the Thurstonian DOD-model proposed:
  - Unequal variance
  - Sequence effects
  - Replications
- Implementation in sensR

# Conclusions

Main results:

- DOD protocol brought up to speed with other discrimination protocols
- Thurstonian model developed
- Three extensions of the Thurstonian DOD-model proposed:
  - Unequal variance
  - Sequence effects
  - Replications
- Implementation in sensR (soon. . . )

# Perspectives

Open questions and future work:

# Perspectives

Open questions and future work:

- How does the DOD protocol compare with Triangle, Tetrad, etc.?

# Perspectives

Open questions and future work:

- How does the DOD protocol compare with Triangle, Tetrad, etc.?
- How many categories should we choose? — power, $Var(d')$.

# Perspectives

Open questions and future work:

- How does the DOD protocol compare with Triangle, Tetrad, etc.?
- How many categories should we choose? — power, $Var(d')$.
- How likely are we to detect unequal variance, sequence effects and heterogeneity?

# Perspectives

Open questions and future work:

- How does the DOD protocol compare with Triangle, Tetrad, etc.?
- How many categories should we choose? — power, $Var(d')$.
- How likely are we to detect unequal variance, sequence effects and heterogeneity?
- How do these effects influence $d'$ and power?

# Perspectives

Open questions and future work:

- How does the DOD protocol compare with Triangle, Tetrad, etc.?
- How many categories should we choose? — power, $Var(d')$.
- How likely are we to detect unequal variance, sequence effects and heterogeneity?
- How do these effects influence $d'$ and power?
- Are we able to distinguish between decision rules for DOD and A-not A with sureness?

Thanks to the scientific committee

Thank you for your attention!

# References

Aust, L. B., M. C. Gacula Jr., S. A. Beard, and R. A. Washam (1985). Degree of difference test method in sensory evaluation of heterogeneous product types. *Journal of Food Science 50*, 511–513.

Bi, J. (2002). Statistical models for the degree of difference test. *Food Quality and Preference 13*, 13–37.

Irwin, R. J., J. A. Stillman, M. J. Hautus, and L. M. Huddleston (1993). The measurement of taste discrimination with the same-different task: a detection-theory analysis. *Journal of Sensory Studies 8*, 229–239.

O'Mahony, M. and B. Rousseau (2002). Discrimination testing: a few ideas, old and new. *Food Quality and Preference 14*, 157–164.