A new statistic to detect segmentation or unequal variance in 2-Alternative Choice (2-AC) testing

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 13th 2012

DTU Informatics
Department of Informatics and Mathematical Modelling

Paired preference testing

2 products :

- A Chocolate bar (standard)
- B Chocolate bar with darker chocolate

2-Alternative Forced Choice (2-AFC):

- Do you prefer A or B?

Prefer A	Prefer B
☐	☐

2-Alternative ~~Forced~~ Choice (2-AC):

- Do you prefer A or B, or do you have no preference?

Prefer A	No Preference	Prefer B
☐	☐	☐

Paired preference with a *no preference* option

Terminology: *No preference* $\sim$ *No difference* $\sim$ Ties

Why allow for a *no preference* option?

- More information and greater resolution in data
- Products may actually be equally liked
- *No preference* counts may support non-inferiority claims

Why avoid a *no preference* option?

- Statistical methods less well-known

Placebo experiments and identicality norms

Consider the data:

	Prefer A	No Preference	Prefer B	Total
All counts	90	20	90	200
Segment 1	8	10	82	100
Segment 2	82	10	8	100

- Are there no differences wrt. preference in the population?
- What if there are two opposing segments?

Ennis and Ennis (2012) suggest:

1. Perform placebo experiment
2. Estimate the *identicality norm*:
   The expected proportion of counts for identical products

Ennis, D. M. and J. M. Ennis (2012). Accounting for no difference/preference responses or ties in choice experiments. *Food Quality and Preference 23*, 13-17.

Example: Comparing data with an identicality norm

Ennis' Approach:

	Prefer A	No Preference	Prefer B	Total
Data	25	15	60	100
Identicality norm	0.4	0.2	0.4	—

$$X_2^2 = (25-40)^2/40 + (15-20)^2/20 + (60-40)^2/40$$
$$= 5.625 + 1.250 + 10.00 = 16.875$$
$$p\text{-value} = 0.00022$$

- Assumes identicality norm known without error
- Uncertainty in the placebo experiment not taken into account!

Example: Comparing data with an identicality norm

How do we take the uncertainty in the placebo experiment into account?

Assume $n = 100$ in placebo experiment:

	Prefer A	No Preference	Prefer B	Total
Data	25	15	60	100
Placebo data	40	20	40	100

Expected counts:

	Prefer A	No Preference	Prefer B	Total
Data	32.5	17.5	50	100
Placebo data	32.5	17.5	50	100

The standard (genuine) Pearson χ^2 test:

$$X_2^2 = (25 - 32.5)^2/32.5 + (40 - 32.5)^2/32.5 + \ldots + (40 - 50)^2/50 = 8.18$$

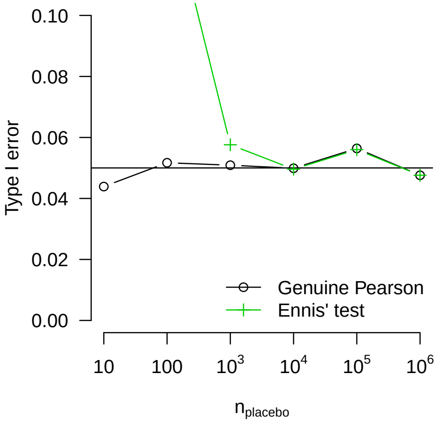
p-value $= 0.0168$ (previous p-value $= 0.00022$)

Effect of sample size in placebo experiment

Standard Pearson test on 2×3 table:

$n_{placebo}$	χ^2_2 statistic	p-value
20	2.80	0.24619
50	5.50	0.06393
100	8.18	0.01677
1000	15.15	0.00051
10^9	16.87	0.00022

Ennis & Ennis (2012):
$X^2 = 16.87$ and p-value $= 0.00022$



Preliminary results and purpose of this work

Preliminary results:

- *No preference* votes contain information
- Don't ignore the uncertainty in the placebo data
- The genuine Pearson test on the 2×3 table is a better option

Are there even better tests?

Purpose of this work: Find a good test for 2-AC testing

Desirable properties of a good test:

- Appropriate type I error
- High power
- Insightful interpretation
- Easy to compute

Approach

1. Consider 5 test statistics
2. Compare the power of the 5 tests in a simulation study

Parameterization and test statistics

Parameterization:

Experiment	Prefer A	No Preference	Prefer B
Placebo	$p_0(1-s_0)$	s_0	$(1-p_0)(1-s_0)$
Preference	$p_1(1-s_1)$	s_1	$(1-p_1)(1-s_1)$

Test statistics:

Test	Null Hypothesis	Alternative Hypothesis	df
Tie effects	$s_0 = s_1$	$s_0 \neq s_1$	1
Directional effects	$p_1 = 0.5$	$p_1 \neq 0.5$	1
Genuine Pearson	$s_0 = s_1$ and $p_0 = p_1$	$s_0 \neq s_1$ or $p_0 \neq p_1$	2
Modified Pearson	$s_0 = s_1$ and $p_1 = 0.5$	$s_0 \neq s_1$ or $p_1 \neq 0.5$	2
Pooled Test	$s_0 = s_1$ and $p_1 = 0.5$	$s_0 \neq s_1$ or $p_1 \neq 0.5$	2

- Note: $p_0 = 0.5$ is given by the design!
- The Genuine Pearson test is NOT the right test

Settings for power simulations

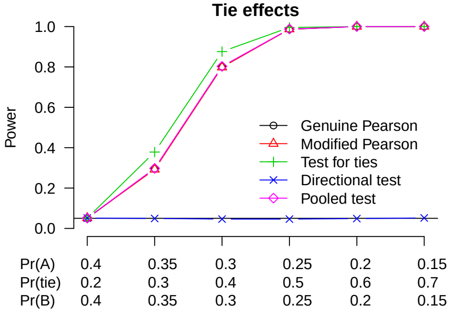
Placebo experiment (true identicality norm):

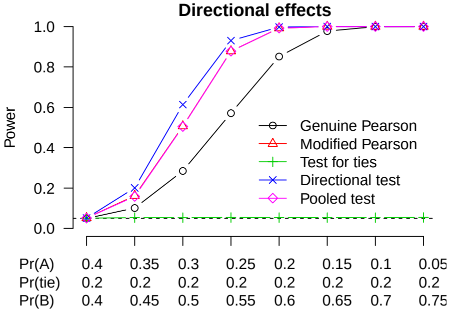
Prefer A	No Preference	Prefer B
0.4	0.2	0.4

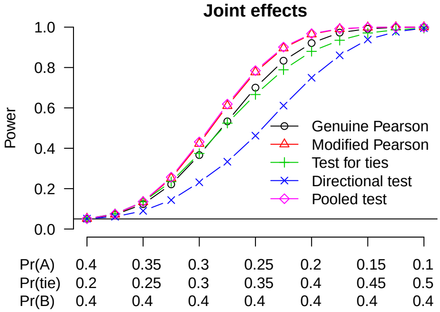
Power simulations in 6 settings:

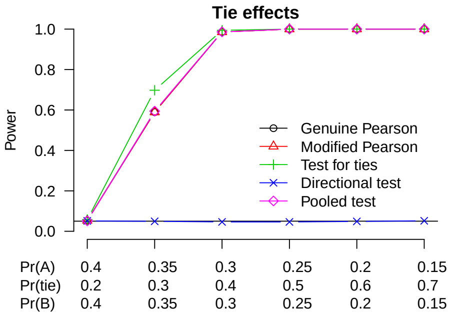
	Structures in preference data		
Placebo sample size	Tie effects	Directional effects	Joint effects
100	1A	1B	1C
1.000.000	2A	2B	2C

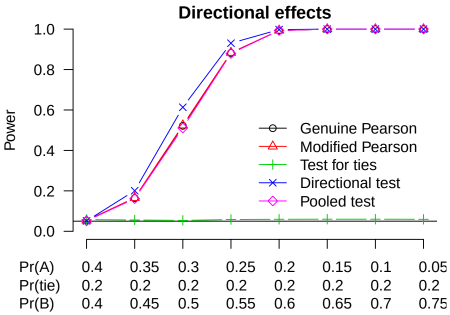
- $n_{preference} = 100$
- 10.000 simulations at each point

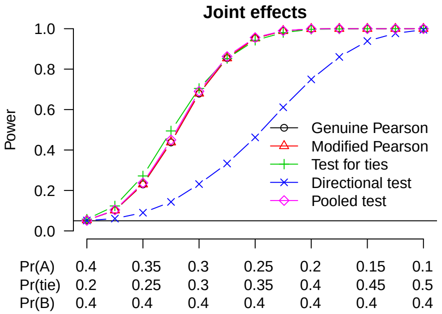

Tie effects
Power
1.0
0.8
0.6
0.4
0.2
0.0
Genuine Pearson
Modified Pearson
Test for ties
Directional test
Pooled test
Pr(A) 0.4 0.35 0.3 0.25 0.2 0.15
Pr(tie) 0.2 0.3 0.4 0.5 0.6 0.7
Pr(B) 0.4 0.35 0.3 0.25 0.2 0.15


Directional effects
Power
1.0
0.8
0.6
0.4
0.2
0.0
Genuine Pearson
Modified Pearson
Test for ties
Directional test
Pooled test
Pr(A) 0.4 0.35 0.3 0.25 0.2 0.15 0.1 0.05
Pr(tie) 0.2 0.2 0.2 0.2 0.2 0.2 0.2 0.2
Pr(B) 0.4 0.45 0.5 0.55 0.6 0.65 0.7 0.75


Joint effects
Power
1.0
0.8
0.6
0.4
0.2
0.0
Genuine Pearson
Modified Pearson
Test for ties
Directional test
Pooled test
Pr(A) 0.4 0.35 0.3 0.25 0.2 0.15 0.1
Pr(tie) 0.2 0.25 0.3 0.35 0.4 0.45 0.5
Pr(B) 0.4 0.4 0.4 0.4 0.4 0.4 0.4


Tie effects
Power
1.0
0.8
0.6
0.4
0.2
0.0
Genuine Pearson
Modified Pearson
Test for ties
Directional test
Pooled test
Pr(A) 0.4 0.35 0.3 0.25 0.2 0.15
Pr(tie) 0.2 0.3 0.4 0.5 0.6 0.7
Pr(B) 0.4 0.35 0.3 0.25 0.2 0.15


Directional effects
Power
1.0
0.8
0.6
0.4
0.2
0.0
Genuine Pearson
Modified Pearson
Test for ties
Directional test
Pooled test
Pr(A) 0.4 0.35 0.3 0.25 0.2 0.15 0.1 0.05
Pr(tie) 0.2 0.2 0.2 0.2 0.2 0.2 0.2 0.2
Pr(B) 0.4 0.45 0.5 0.55 0.6 0.65 0.7 0.75


Joint effects
Power
1.0
0.8
0.6
0.4
0.2
0.0
Genuine Pearson
Modified Pearson
Test for ties
Directional test
Pooled test
Pr(A) 0.4 0.35 0.3 0.25 0.2 0.15 0.1
Pr(tie) 0.2 0.25 0.3 0.35 0.4 0.45 0.5
Pr(B) 0.4 0.4 0.4 0.4 0.4 0.4 0.4

Example — new insights

Example data:

	Prefer A	No Preference	Prefer B	Total
Placebo exp.	81	45	74	200
Preference exp.	37	12	51	100

ANOVA-like analysis:

Test	χ^2	df	p-value
Pooled test	7.00	2	0.030
Tie effects	4.78	1	0.029
Directional effects	2.23	1	0.136
Modified Pearson	7.20	2	0.027

Final points

Conclusions and recommendations:

- Placebo data contain valuable information
- Don't ignore the uncertainty in the placebo data
- The modified Pearson and Pooled statistics have the highest power against general alternatives
- Use the Pooled statistic to provide insight into the structure of the data

Open questions:

- What may cause tie-effects?
  - Segmentation
  - Heterogeneity in preference
  - Unequal variances in the underlying perceptual distributions

A new statistic to detect segmentation or unequal variance in 2-Alternative Choice (2-AC) testing

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 13th 2012

DTU Informatics
Department of Informatics and Mathematical Modelling