Handling Ties

Analysis of Ties in Input and Output Data of Rankings

TECHNISCHE
UNIVERSITÄT
DARMSTADT

Outline

TECHNISCHE UNIVERSITÄT DARMSTADT

Outline

Genuine Input Ties

Incorporating Ties

Perturbation Analysis

Real Data Sets

Induced Input Ties

Output Ties

Summary

- Genuine Input Ties
- Incorporating Ties
- Perturbation Analysis
- Real Data Sets
- Induced Input Ties
- Output Ties
- Summary

Genuine Input Ties

TECHNISCHE UNIVERSITÄT DARMSTADT

Outline
Genuine Input Ties
Incorporating Ties
Perturbation Analysis
Real Data Sets
Induced Input Ties
Output Ties
Summary

- Tied results in the input data
- Frequency depends on
  - data source
  - tie resolution policy

Incorporating Ties

TECHNISCHE UNIVERSITÄT DARMSTADT

Outline

Genuine Input Ties

Incorporating Ties

Colley's Method

Massey's Method

Markov's Method

Elo's, Keener's, OD Method

Perturbation Analysis

Real Data Sets

Induced Input Ties

Output Ties

Summary

- Colley's method does not account for ties
- Markov's methods depends on voting mechanism used
- Elo's, Keener's, Massey's and OD method account for ties

Incorporating Ties

TECHNISCHE UNIVERSITÄT DARMSTADT

Outline

Genuine Input Ties

Incorporating Ties

Colley's Method

Massey's Method

Markov's Method

Elo's, Keener's, OD Method

Perturbation Analysis

Real Data Sets

Induced Input Ties

Output Ties

Summary

To compare methods with ties and methods without we need to

- derive a Colley's method accounting for ties
- create a Massey's method ignoring ties
- choose a Markov's method allowing for both
- modify Elo's, Keener's and the OD method

Colley's Method

TECHNISCHE UNIVERSITÄT DARMSTADT

Outline

Genuine Input Ties

Incorporating Ties

Colley's Method

Massey's Method

Markov's Method

Elo's, Keener's, OD Method

Perturbation Analysis

Real Data Sets

Induced Input Ties

Output Ties

Summary

$C * r = b$, with

- $C_{n \times n}$ having entries as follows

$$C_{ij} = \begin{cases} 2 + t_i, & \text{i=j} \\ -n_{ij}, & \text{i}\neq\text{j} \end{cases}$$

- $r_{n \times 1}$ being the unknown Colley rating vector
- $b_{n \times 1}$ defined as $b_i = 1 + \frac{1}{2}(w_i - l_i)$

Colley's Method

Outline
Genuine Input Ties
Incorporating Ties
Colley's Method
Massey's Method
Markov's Method
Elo's, Keener's, OD Method
Perturbation Analysis
Real Data Sets
Induced Input Ties
Output Ties
Summary

Ties in Colley's method

- are ignored
- represent an equal chance for either team winning - or losing
- can be emulated by creating two artificial games
- do not alter vector b
- increment C_{ij} by 1 and decrement C_{ji} by 1
- thus preserving the Colley property

$$\sum_{i=0}^{n} r_i = \frac{n}{2}$$

Massey's Method

$M * r = p$, with

- $M_{n \times n}$ having entries as follows

$$M_{ij} = \begin{cases} t_i, & \text{i=j} \\ -n_{ij}, & \text{i}\neq\text{j} \end{cases}$$

- $r_{n \times 1}$ being the unknown Massey rating vector
- $p_{n \times 1}$ being the vector of all teams' point differentials

Massey's Method

TECHNISCHE UNIVERSITÄT DARMSTADT

Outline

Genuine Input Ties

Incorporating Ties

Colley's Method

Massey's Method

Markov's Method

Elo's, Keener's, OD Method

Perturbation Analysis

Real Data Sets

Induced Input Ties

Output Ties

Summary

Ties in Massey's method

- are naturally accounted for
- increment M_{ij} and M_{ji}
- do not change p
- can be ignored when forming M to create a No-Ties Method

Markov's Method

Outline

Genuine Input Ties

Incorporating Ties

Colley's Method

Massey's Method

Markov's Method

Elo's, Keener's, OD Method

Perturbation Analysis

Real Data Sets

Induced Input Ties

Output Ties

Summary

Standard Markov voting procedures:

1. Loser casts one vote for each team lost against.
2. Loser casts one vote for each point lost to the other team.
3. Loser and winner cast one vote for each point lost to one another.

Markov's Method

TECHNISCHE UNIVERSITÄT DARMSTADT

Outline
Genuine Input Ties
Incorporating Ties
Colley's Method
Massey's Method
Markov's Method
Elo's, Keener's, OD Method
Perturbation Analysis
Real Data Sets
Induced Input Ties
Output Ties
Summary

In the context of ties:

1. Tied teams cast half a vote each for the other team.
2. Loser casts one vote for each point lost on average.
3. Ignore tied events for a no-ties variant.

Elo's, Keener's, OD Method

TECHNISCHE UNIVERSITÄT DARMSTADT

Outline

Genuine Input Ties

Incorporating Ties

Colley's Method

Massey's Method

Markov's Method

Elo's, Keener's, OD Method

Perturbation Analysis

Real Data Sets

Induced Input Ties

Output Ties

Summary

Ties in Elo's Method

- are explicitly taken care of
- can be ignored for a no-ties method

Elo's, Keener's, OD Method

TECHNISCHE UNIVERSITÄT DARMSTADT

Outline

Genuine Input Ties

Incorporating Ties

Colley's Method

Massey's Method

Markov's Method

Elo's, Keener's, OD Method

Perturbation Analysis

Real Data Sets

Induced Input Ties

Output Ties

Summary

Ties in Keener's and the OD Method

- are naturally accounted for
- can be excluded by setting tied scores to 0

Perturbation Analysis

TECHNISCHE UNIVERSITÄT DARMSTADT

Outline

Genuine Input Ties

Incorporating Ties

Perturbation Analysis

Real Data Sets

Induced Input Ties

Output Ties

Summary

- Apply variants with and without ties to a data set devoid of ties
- Introduce a single tie into data set
- Compare the rankings produced by the variant methods

Perturbation Analysis

TECHNISCHE UNIVERSITÄT DARMSTADT

Outline
Genuine Input Ties
Incorporating Ties
Perturbation Analysis
Real Data Sets
Induced Input Ties
Output Ties
Summary

Let

- C be the Colley matrix for the no-ties variant
- $\widetilde{C}$ denote the Colley matrix for the variant including ties
- e denote a tied input
- r be the rating vector related to C
- $\tilde{r}$ be the rating vector related to $\widetilde{C}$

Perturbation Analysis

$$\widetilde{C} = C + (e_i - e_j)(e_i - e_j)^T$$

$$\tilde{r} = r - \left(\frac{r_i - r_j}{1 + [C^{-1}]_{ii} - 2[C^{-1}]_{ij} + [C^{-1}]_{jj}}\right) C^{-1}(e_i - e_j)$$

Perturbation Analysis

TECHNISCHE UNIVERSITÄT DARMSTADT

Handling Ties

Manuel Weidmann (Speaker), Prof. Dr. Johannes Fürnkranz

Outline

Genuine Input Ties

Incorporating Ties

Perturbation Analysis

Real Data Sets

Induced Input Ties

Output Ties

Summary

With $\epsilon = r_i - r_{i+1}, \epsilon > 0$ denoting the difference in pre-disturbance ratings this implies

$$\tilde{r}_i < \tilde{r}_{i+1} \iff$$

$$\epsilon < \left(\frac{(r_i - r_j)([C^{-1}]_{ii} - [C^{-1}]_{ij} - [C^{-1}]_{i+1,i} + [C^{-1}]_{i+1,j})}{1 + [C^{-1}]_{ii} - 2[C^{-1}]_{ij} + [C^{-1}]_{jj}} * C^{-1}(e_i - e_j)\right)$$

Let $r_i < r_j$:

- if $r_i \approx r_j$ teams i and j are unlikely to change in rank
- if $r_i \gg r_j$ team i is likely to drop in rank

Real Data Sets

Outline

Genuine Input Ties

Incorporating Ties

Perturbation Analysis

Real Data Sets

Movies

NHL Hockey Teams

Induced Input Ties

Output Ties

Summary

Movies

Outline
Genuine Input Ties
Incorporating Ties
Perturbation Analysis
Real Data Sets
Movies
NHL Hockey Teams
Induced Input Ties
Output Ties
Summary

Rank	Ties Ignored	Ties Incorporated	Change
1	Casablanca	Casablanca	-
2	Schindler's List	Schindler's List	-
3	Shawshank Redemption	Rear Window	1 up
4	Rear Window	Shawshank Redemption	1 down
5	The Godfather	The Godfather	-
6	Citizen Kane	To Kill a Mockingbird	2 up
8	To Kill a Mockingbird	Citizen Kane	2 down
11	Pulp Fiction	Raging Bull	3 up
12	It's a Wonderful Life	It's a Wonderful Life	-
13	Taxi Driver	Taxi Driver	-
14	Raging Bull	Pulp Fiction	3 down

NHL Hockey Teams

TECHNISCHE UNIVERSITÄT DARMSTADT

Handling Ties

Manuel Weidmann (Speaker), Prof. Dr. Johannes Fürnkranz

Outline

Genuine Input Ties

Incorporating Ties

Perturbation Analysis

Real Data Sets

Movies

NHL Hockey Teams

Induced Input Ties

Output Ties

Summary

Induced Input Ties

TECHNISCHE UNIVERSITÄT DARMSTADT

Outline

Genuine Input Ties

Incorporating Ties

Perturbation Analysis

Real Data Sets

Induced Input Ties

Motivation

Methods

Analysis

Output Ties

Summary

- Modifications of input data
- Change non-tied events to a tie

Induced Input Ties

Outline
Genuine Input Ties
Incorporating Ties
Perturbation Analysis
Real Data Sets
Induced Input Ties
Motivation
Methods
Analysis
Output Ties
Summary

Ties are often broken

- seemingly at random
- irrespective of the teams' actual performance
- after regular match time

Induced Input Ties

TECHNISCHE UNIVERSITÄT DARMSTADT

Handling Ties

Manuel Weidmann (Speaker), Prof. Dr. Johannes Fürnkranz

Outline

Genuine Input Ties

Incorporating Ties

Perturbation Analysis

Real Data Sets

Induced Input Ties

- Motivation
- Methods
- Analysis

Output Ties

Summary

Costa Rica	1 – 1 (5 – 3)	Greece

Scoring summary

Bryan Ruiz (#10) 52' ⚽ Socratis Papastathopoulos (#19) 90+1'

Team stats

6	Total shots	24
2	Shots on target	13
23	Fouls	16
43%	Possession (%)	57%
3	Corners	11
7	Saves	0
0	Offsides	10

Induced Input Ties

TECHNISCHE UNIVERSITÄT DARMSTADT

Outline
Genuine Input Ties
Incorporating Ties
Perturbation Analysis
Real Data Sets
Induced Input Ties
Motivation
Methods
Analysis
Output Ties
Summary

Final results may

- fail to portray the teams' actual strength
- create a false sense of precision
- skew the ranking
- impede a ranking's predictive capabilities

Methods

TECHNISCHE UNIVERSITÄT DARMSTADT

Outline
Genuine Input Ties
Incorporating Ties
Perturbation Analysis
Real Data Sets
Induced Input Ties
Motivation
Methods
Analysis
Output Ties
Summary

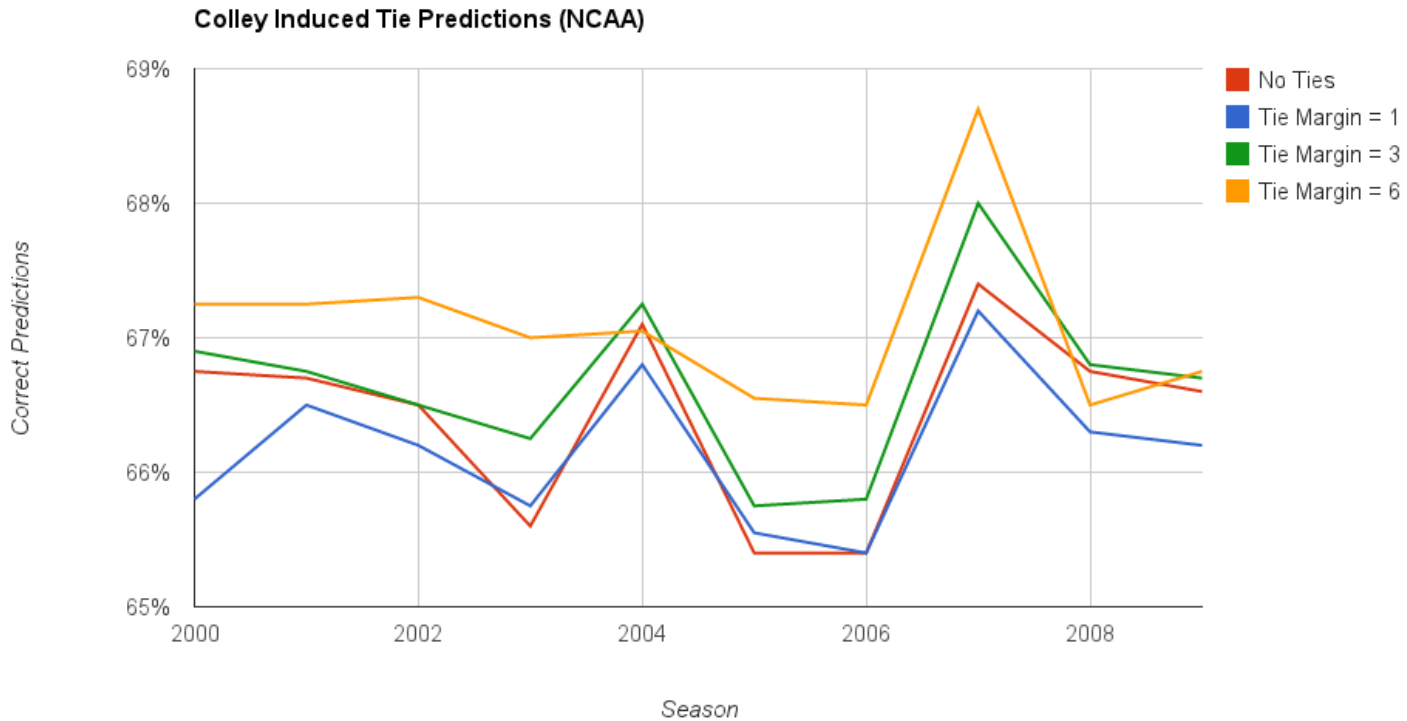
Induce a tie if

- a winner is only determined after regular play time
- points differ only by a small margin
- match statistics indicate comparable performance

Analysis

TECHNISCHE UNIVERSITÄT DARMSTADT

Outline
Genuine Input Ties
Incorporating Ties
Perturbation Analysis
Real Data Sets
Induced Input Ties
Motivation
Methods
Analysis
Output Ties
Summary



Output Ties

Outline

Genuine Input Ties

Incorporating Ties

Perturbation Analysis

Real Data Sets

Induced Input Ties

Output Ties

- Recapitulation
- Resolution Methods
- Ramifications

Summary

Recapitulation

TECHNISCHE UNIVERSITÄT DARMSTADT

Outline
Genuine Input Ties
Incorporating Ties
Perturbation Analysis
Real Data Sets
Induced Input Ties
Output Ties
Recapitulation
Resolution Methods
Ramifications
Summary

Rankings are total preorders, i.e. relations on a set S that are

- total $(\forall x, y \in S : x \leq y \vee y \leq x)$
- transitive $(\forall x, y, z \in S : x \leq y \wedge y \leq z \Rightarrow x \leq z)$
- but not antisymmetric $(\exists x, y \in S : x \leq y \wedge y \leq x \wedge x \neq y)$

Resolution Methods

Outline
Genuine Input Ties
Incorporating Ties
Perturbation Analysis
Real Data Sets
Induced Input Ties
Output Ties
Recapitulation
Resolution Methods
Ramifications
Summary

- Standard Competition Ranking ("1 2 2 4")
- Modified Competition Ranking ("1 3 3 4")
- Dense Ranking ("1 2 2 3")
- Ordinal Ranking ("1 2 3 4")
- Fractional Ranking ("1 2.5 2.5 4")

Ramifications

TECHNISCHE UNIVERSITÄT DARMSTADT

Outline

Genuine Input Ties

Incorporating Ties

Perturbation Analysis

Real Data Sets

Induced Input Ties

Output Ties

Recapitulation

Resolution Methods

Ramifications

Summary

- Psychological Effects
- Statistical Effects

Statistical Effects - Rank Sum Test

TECHNISCHE UNIVERSITÄT DARMSTADT

Handling Ties

Manuel Weidmann (Speaker), Prof. Dr. Johannes Fürnkranz

Outline

Genuine Input Ties

Incorporating Ties

Perturbation Analysis

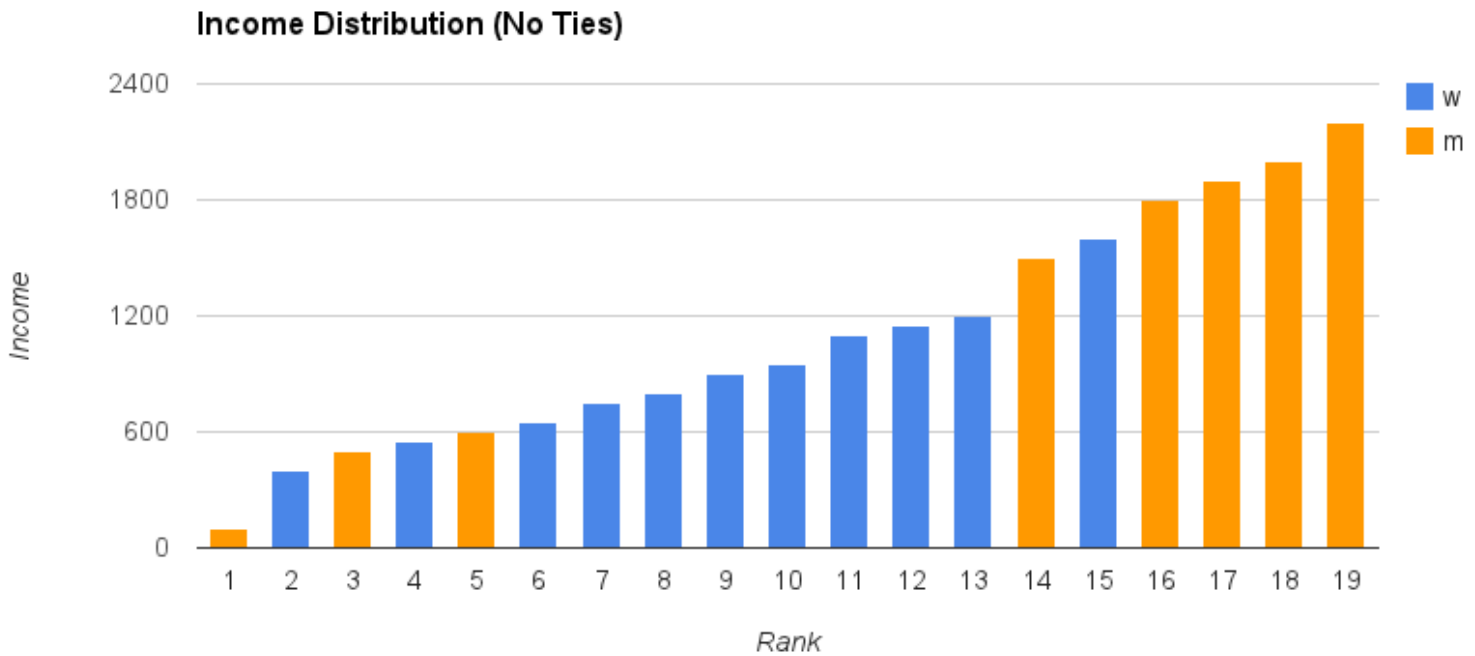
Real Data Sets

Induced Input Ties

Output Ties

- Recapitulation
- Resolution Methods
- Ramifications

Summary



Statistical Effects - Rank Sum Test

TECHNISCHE UNIVERSITÄT DARMSTADT

Handling Ties

Manuel Weidmann (Speaker), Prof. Dr. Johannes Fürnkranz

Outline

Genuine Input Ties

Incorporating Ties

Perturbation Analysis

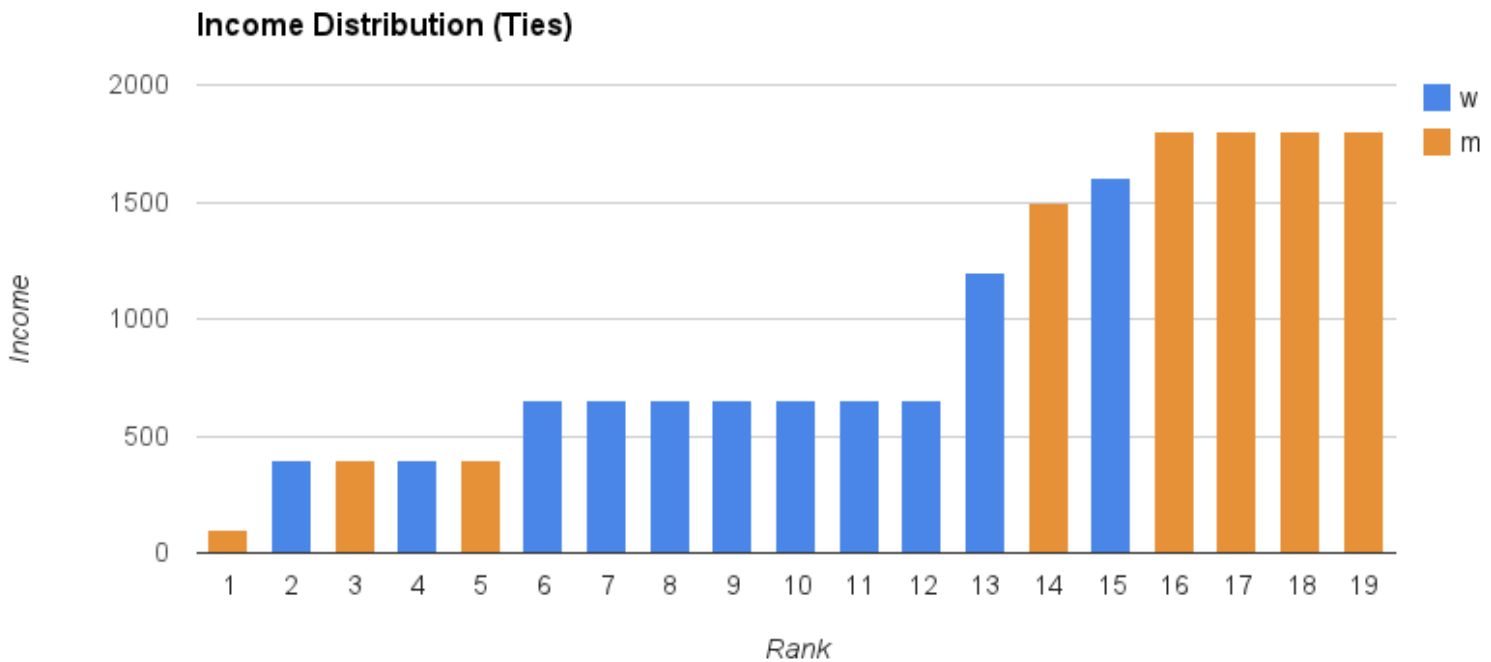
Real Data Sets

Induced Input Ties

Output Ties

- Recapitulation
- Resolution Methods
- Ramifications

Summary



Statistical Effects - Rank Sum Test

TECHNISCHE UNIVERSITÄT DARMSTADT

Handling Ties

Manuel Weidmann (Speaker), Prof. Dr. Johannes Fürnkranz

Outline

Genuine Input Ties

Incorporating Ties

Perturbation Analysis

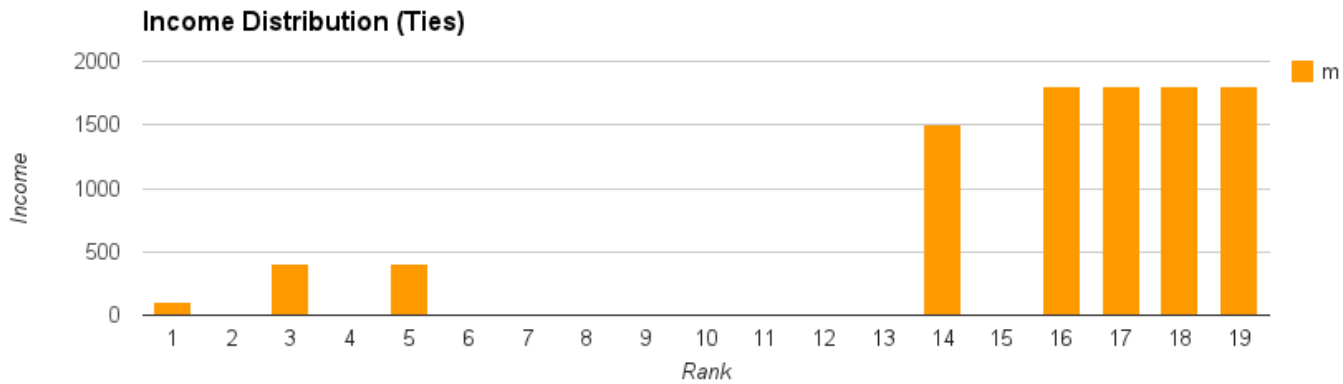
Real Data Sets

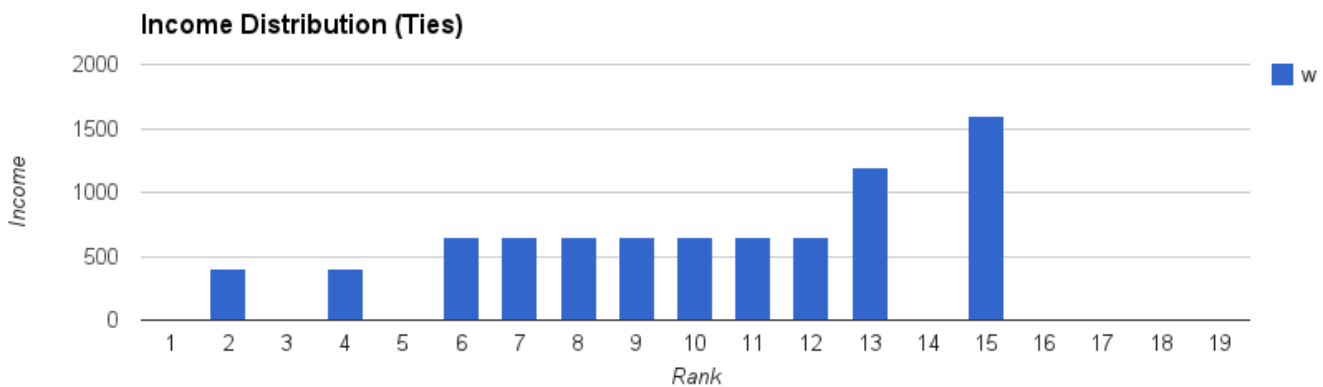
Induced Input Ties

Output Ties

- Recapitulation
- Resolution Methods
- Ramifications

Summary





Statistical Effects - Rank Sum Test

TECHNISCHE UNIVERSITÄT DARMSTADT

Handling Ties

Manuel Weidmann (Speaker), Prof. Dr. Johannes Fürnkranz

Outline

Genuine Input Ties

Incorporating Ties

Perturbation Analysis

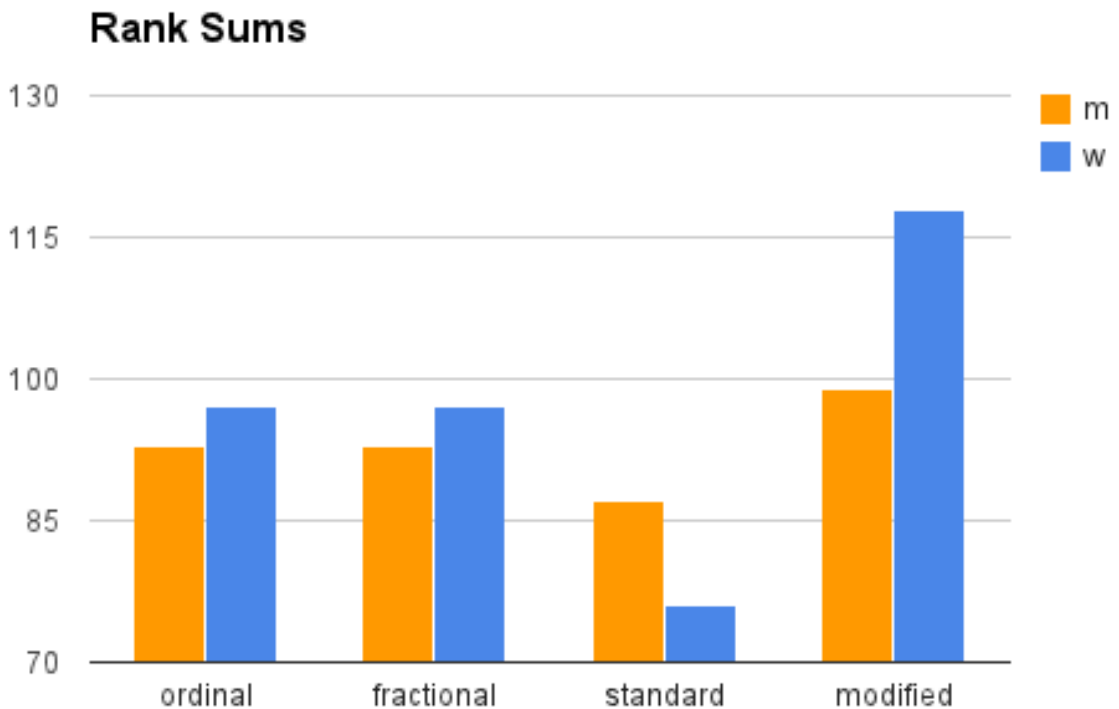
Real Data Sets

Induced Input Ties

Output Ties

- Recapitulation
- Resolution Methods
- Ramifications

Summary



Summary

Outline
Genuine Input Ties
Incorporating Ties
Perturbation Analysis
Real Data Sets
Induced Input Ties
Output Ties
Summary

- Accounting for ties is easy
- Input ties influence ranking order
- Inducing ties is beneficial
- Output ties require special care