8th International Conference on Intelligent Data Engineering and Automated Learning (IDEAL'07)

Joint Cutoff Probabilistic Estimation Using Simulation: A Mailing Campaign Application

Antonio Bella
Cèsar Ferri
José Hernández-Orallo
María José Ramírez-Quintana

Technical University of Valencia, Spain

Outline

1. Introduction
2. Campaign design with one product
3. Using simulation and data mining for a campaign design with more than one product
4. Experiments with N products
5. Conclusions

Outline

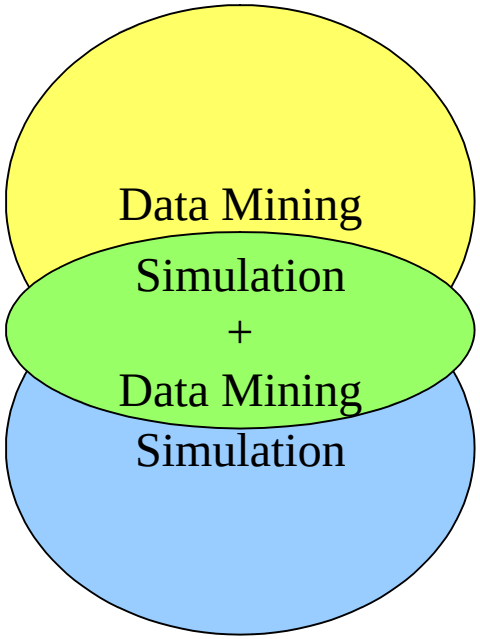
1. Introduction
2. Campaign design with one product
3. Using simulation and data mining for a campaign design with more than one product
4. Experiments with N products
5. Conclusions

1. Introduction (I)

- Example:

Probability	New Coach DM Model	Old Coach DM Model
Trip 1	0.86	0.42
Trip 2	0.56	0.04
Trip 3	0.77	0.21
Trip 4	0.91	0.13

1. Introduction (II)



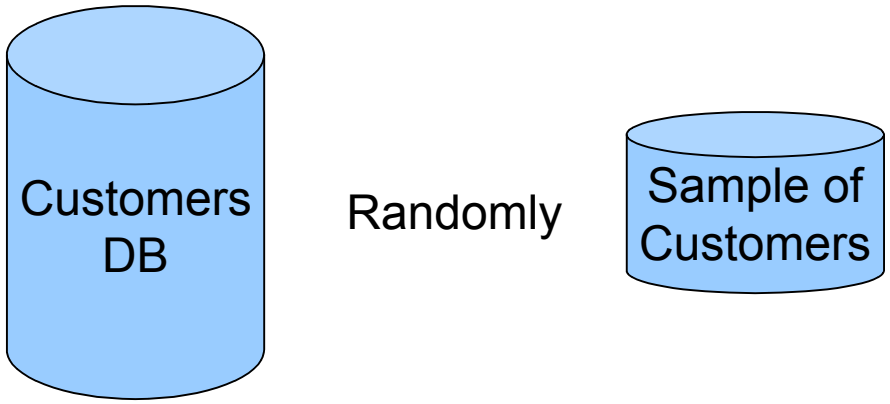
1. Introduction (III)

- DM models accompanied by probabilities
- Train & Test as usual
- Probabilities and constraints are used to estimate the cutoff by simulation
- Simulation framework: basic Petri nets
- A direct-marketing campaign design

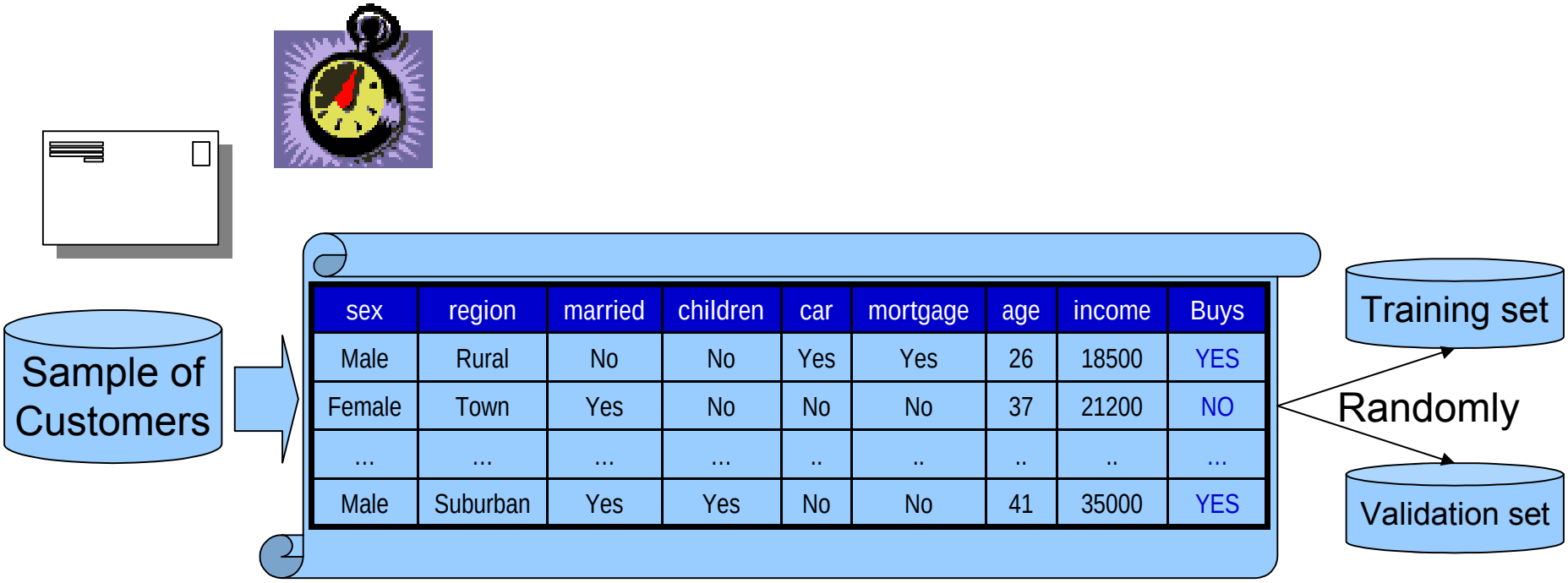
Outline

1. Introduction
2. Campaign design with one product
3. Using simulation and data mining for a campaign design with more than one product
4. Experiments with N products
5. Conclusions

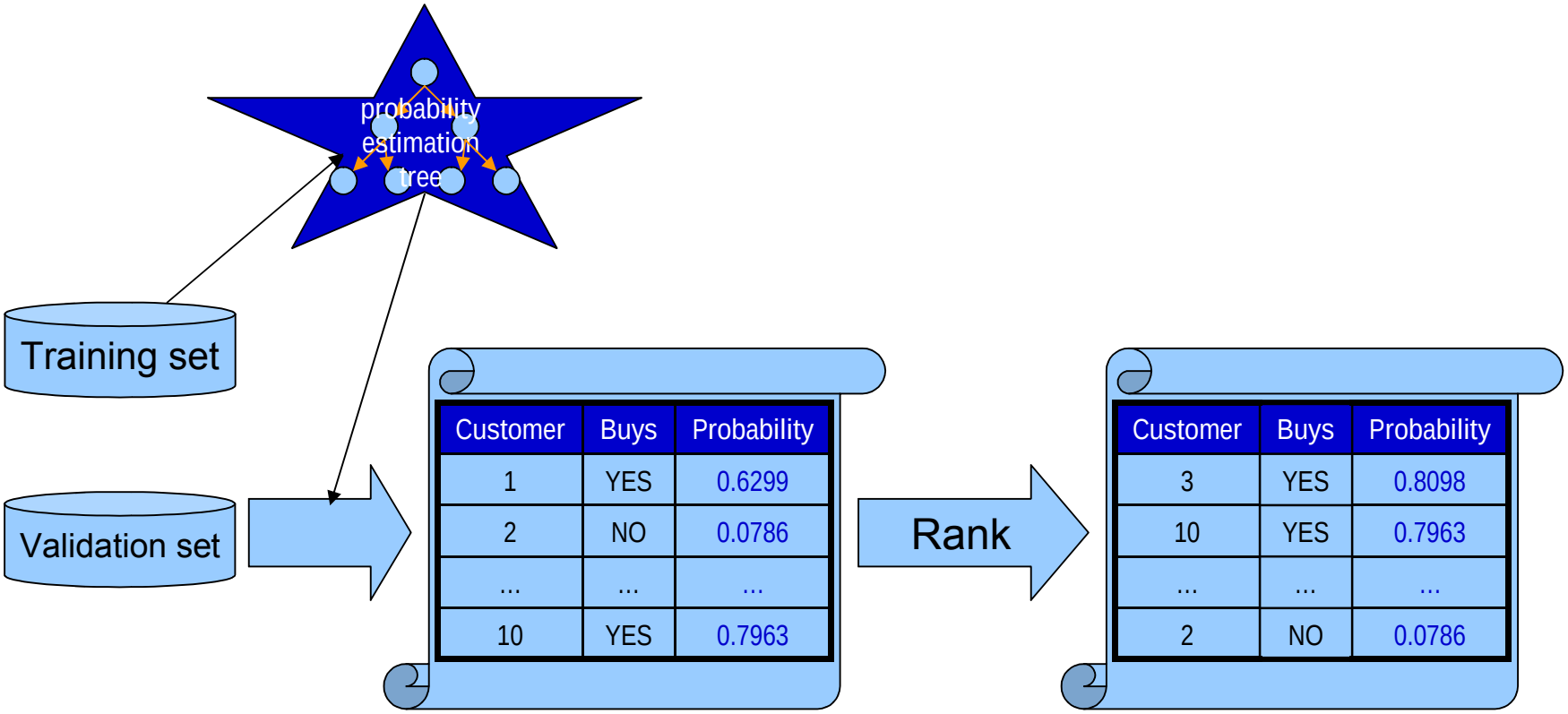
2. Campaign design with one product (I)



2. Campaign design with one product (I)



2. Campaign design with one product (I)



Customer	Buys	Probability
1	YES	0.6299
2	NO	0.0786
...	...	...
10	YES	0.7963

Customer	Buys	Probability
3	YES	0.8098
10	YES	0.7963
...	...	...
2	NO	0.0786

2. Campaign design with one product (II)

- Select the best cutoff
- Additional information:
  - I_{cost}: investment cost
  - b: benefit from selling one product
  - cost: sending cost of a mail
  - C: list of customers ($c_k \in C$)
  - $p(c_k)$: estimated probability
- Accumulated Expected Benefit:

$$- I_{cost} + \Sigma_{k=1..j} (\underbrace{b * p(c_k) - cost}_{\mathbf{E}(\text{Benefit})})$$

- Validation set:
  - real Accumulated Benefit

$$f(c_k) \begin{cases} 0 \\ 1 \end{cases}$$

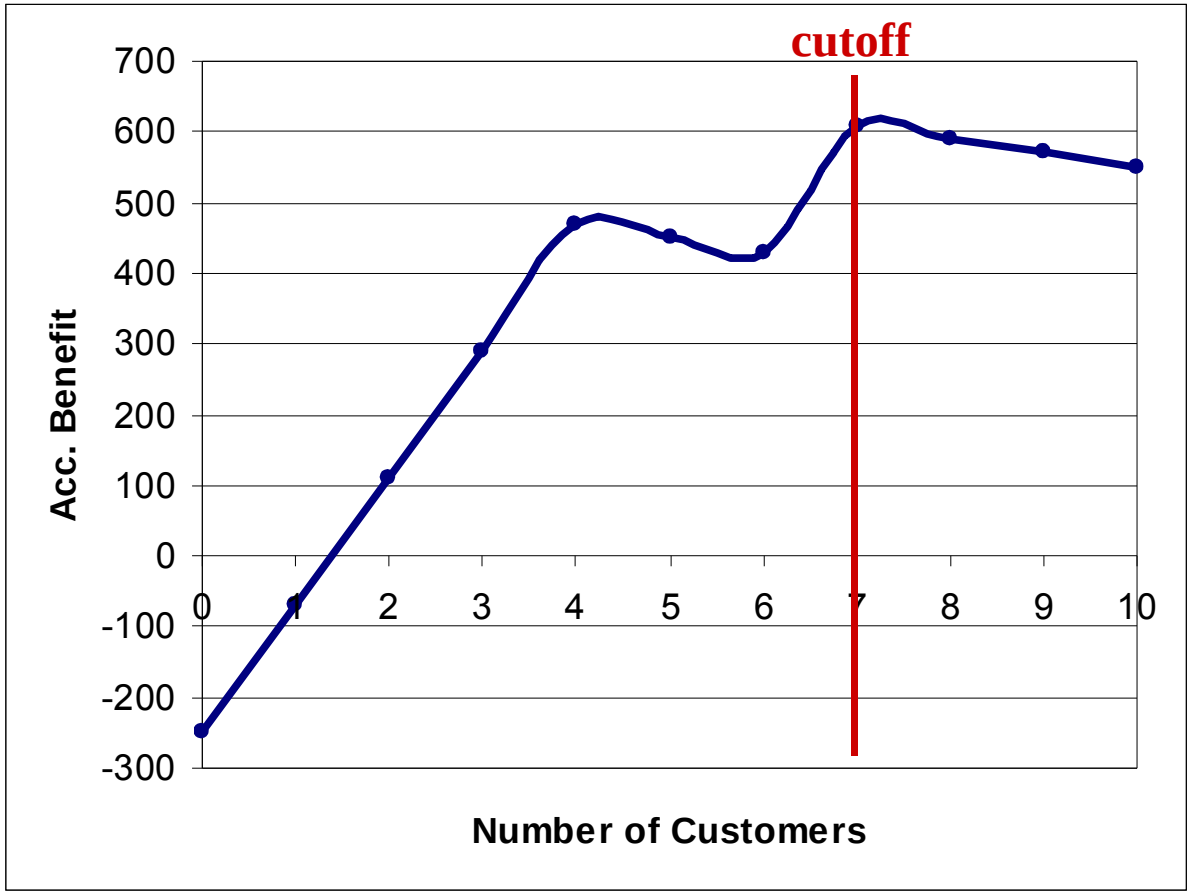
2. Campaign design with one product (III)

- Example:
  - I_{cost} = 250 m.u.
  - b = 200 m.u.
  - cost = 20 m.u.

Customer	Buys	Probability	E(Benefit)	Acc. Benefit
				-250
3	YES	0.8098	141.96	-70
10	YES	0.7963	139.26	110
8	YES	0.6605	112.10	290
1	YES	0.6299	105.98	470
4	NO	0.5743	94.86	450
6	NO	0.5343	86.85	430
5	YES	0.4497	69.94	610
7	NO	0.2675	33.50	590
9	NO	0.2262	24.24	570
2	NO	0.0786	-4.29	550

2. Campaign design with one product (III)

- Example:
  - I_{cost} = 250 m.u.
  - b = 200 m.u.
  - cost = 20 m.u.



Outline

1. Introduction
2. Campaign design with one product
3. Using simulation and data mining for a campaign design with more than one product
4. Experiments with N products
5. Conclusions

3. Using simulation and data mining for a campaign design with more than one product (I)

- Constraints and settings:
  - Stock limits
  - Different benefit for each product
  - Alternative products
  - The same sending cost

3. Using simulation and data mining for a campaign design with more than one product (II)

- Single Approach:
  1. calculate local cutoffs
  2. order all the pairs (customer, product)
  3. global cutoff is the average of local cutoffs
- Joint Simulation Approach:
  1. order all the pairs (customer, product)
  2. calculate best global cuttoff by simulation

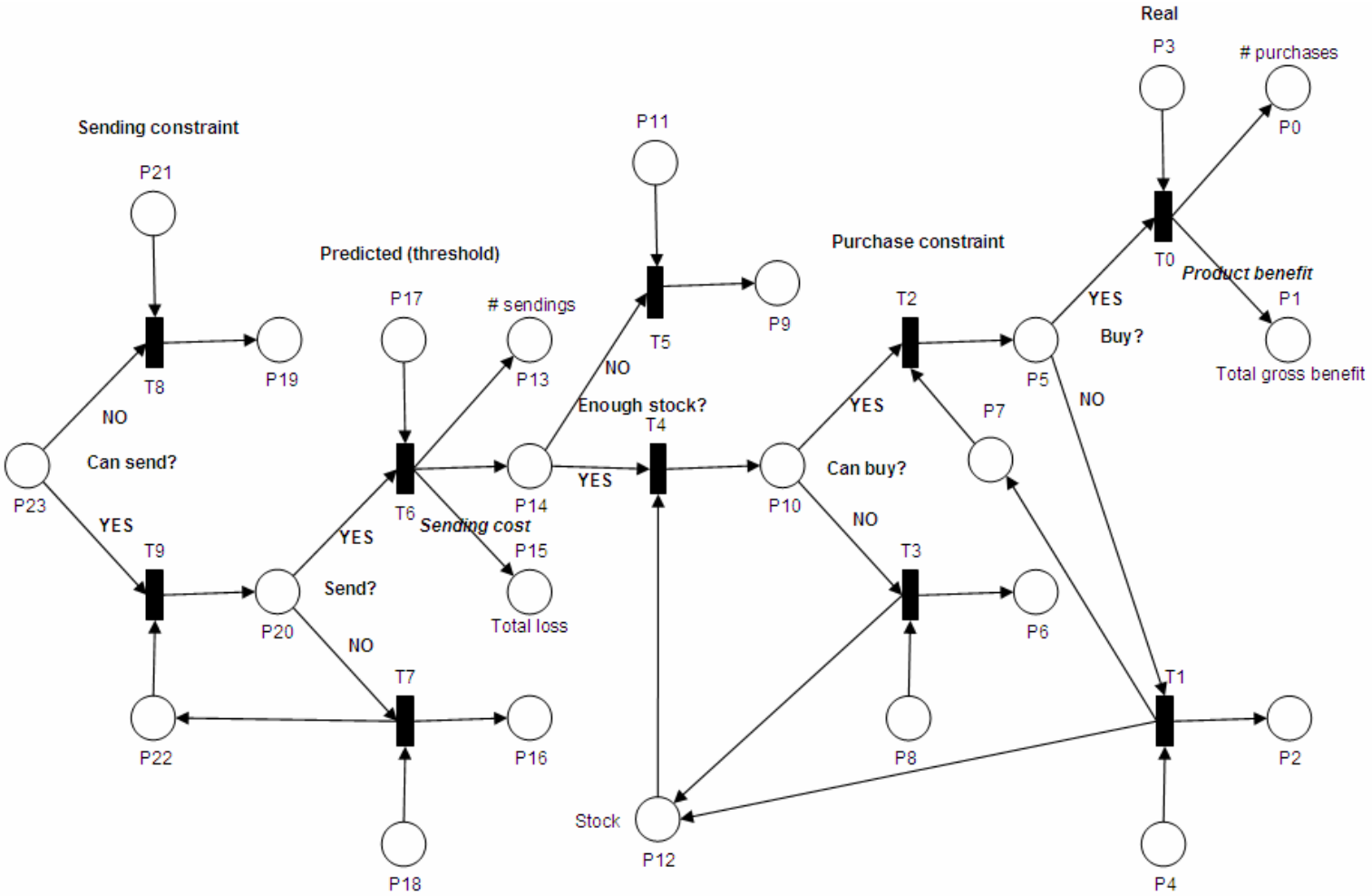
- Example:
  - 10 customers
  - 2 products
  - $Icost_{p1}$ = 150 m.u.
  - $Icost_{p2}$ = 250 m.u.
  - b_1 = 100 m.u.
  - b_2 = 200 m.u.
  - cost = 20 m.u.

Product p_1			
Customer	E(Benefit)	f_{p1}	Acc. Benefit
			-150
2	76.61	1	-70
8	75.71	1	10
9	60.37	0	-10
5	48.19	1	70
1	44.96	1	150
7	30.96	0	130
10	24.58	1	210
3	23.04	0	190
6	7.81	1	270
4	-4.36	0	250

Product p_2			
Customer	E(Benefit)	f_{p2}	Acc. Benefit
			-250
3	141.96	1	-70
10	139.26	1	110
8	112.10	1	290
1	105.98	1	470
4	94.86	0	450
6	86.85	0	430
5	69.94	1	610
7	33.50	0	590
9	25.24	0	570
2	-4.29	0	550

Single & Joint Approaches			
Customer	Product	Acc. Benefit	
		-400	
3	p_2	-220	
10	p_2	-40	
8	p_2	140	
1	p_2	320	
4	p_2	300	
6	p_2	280	
2	p_1	360	
8	p_1	340	
5	p_2	520	Joint
9	p_1	500	
5	p_1	480	
1	p_1	460	
7	p_2	440	
7	p_1	420	
9	p_2	400	
10	p_1	380	Single
3	p_1	360	
6	p_1	440	
2	p_2	420	
4	p_1	400	

3. Using simulation and data mining for a campaign design with more than one product (IV)



Outline

1. Introduction
2. Campaign design with one product
3. Using simulation and data mining for a campaign design with more than one product
4. Experiments with N products
5. Conclusions

4. Experiments with N products (I)

- Reference customers file
- Parameters:
  - Number of customers: 10000 (60% training, 20% validation and 20% testing)
  - Number of products: 2, 3 and 4
  - Probability of buying each product: 0.01, 0.05, 0.2, 0.5, 0.8, 0.95 and 0.99
  - Benefits for each product: 100 m.u. for the first product and 100, 200, 500 or 1000 m.u. for the others
  - Sending cost (the same for all products): 10, 20, 50 or 90 m.u.
  - Stock for each product: 0.1, 0.2, 0.5 or 1 (multiplied by number of customers)
  - Investment cost for each product: benefits of the product multiplied by stock of the product and divided by 20
  - Correlation (how similar the products are): 0.25, 0.5, 0.75 or 1

4. Experiments with N products (II)

	2 products		3 products		4 products	
	Single	Joint	Single	Joint	Single	Joint
Benefits	165626	171225	182444	186805	220264	231771
Single	-	V	-	V	-	V
Joint	X	-	X	-	X	-

- Friedman test: wins (V) / loses (X) ($\alpha = 0.005$)
- Average of many different situations and parameters

Outline

1. Introduction
2. Campaign design with one product
3. Using simulation and data mining for a campaign design with more than one product
4. Experiments with N products
5. Conclusions

5. Conclusions

- New framework to combine simulation and data mining to address decision making problems
- Simulation-based method outperforms classical analytical one
- Future work:
  - Other kind of problems
  - Probability calibration

Thanks for your attention!

Antonio Bella

http://www.dsic.upv.es/~abella

abella@dsic.upv.es