A column generation approach for correlated random vector generation

Andrés L. Medaglia
Industrial Engineering Department
Centro para la Optimización y Probabilidad Aplicada (COPA)
Universidad de los Andes

Joint work with:
Oscar Guaje, Universidad de los Andes
Jorge Sefair, Arizona State University

Búzios, RJ, Brazil, May 23, 2016
Last updated: May 20, 2016

Universidad de los Andes

Outline

- Correlated random vector generation
- MIP model
- Column generation
- Concluding remarks
- Future work
- References

Universidad de los Andes

Outline

- Correlated random vector generation
- MIP model
- Column generation
- Concluding remarks
- Future work
- References

Correlated random vector generation: Definition

k random variables

$$\mathbf{X} = \begin{bmatrix} x_1(1) & x_2(1) & & x_k(1) \\ x_1(2) & x_2(2) & & x_k(2) \\ \vdots & \vdots & \cdots & \vdots \\ x_1(n) & x_2(k) & & x_k(n) \end{bmatrix}$$

n observations

$$\bar{\boldsymbol{\rho}} = \begin{bmatrix} 1 & \rho_{X_1X_2} & & \rho_{X_1X_k} \\ & 1 & & \rho_{X_2X_k} \\ \vdots & \vdots & \cdots & \vdots \\ \rho_{X_kX_1} & \cdots & & 1 \end{bmatrix}$$

Target correlations

Correlated random vector generation: Some definitions

- Spearman rank correlation:

$$\rho^S_{X_1X_2} = 1 - \frac{6\sum_{i=1}^{n}\left(r_{X_1}(x_1(i)) - r_{X_2}(x_2(i))\right)^2}{n(n^2-1)}$$

- Pearson correlation:

$$\rho^P_{X_1X_2} = \frac{\sum_{i=1}^{n}(x_1(i)-\bar{x}_1)(x_2(i)-\bar{x}_2)}{\sqrt{\sum_{i=1}^{n}(x_1(i)-\bar{x}_1)^2\sum_{i=1}^{n}(x_2(i)-\bar{x}_2)^2}}$$

- Kendall's coefficient of concordance

$$\rho^P_{X_1X_2} = \frac{3\sum_{i=1}^{n}(x_1(i)+x_2(i))^2 - 3n(n+1)^2}{n(n^2-1)}$$

Correlated random vector generation: Some definitions

- Phi coefficient of correlation (for categorical variables):

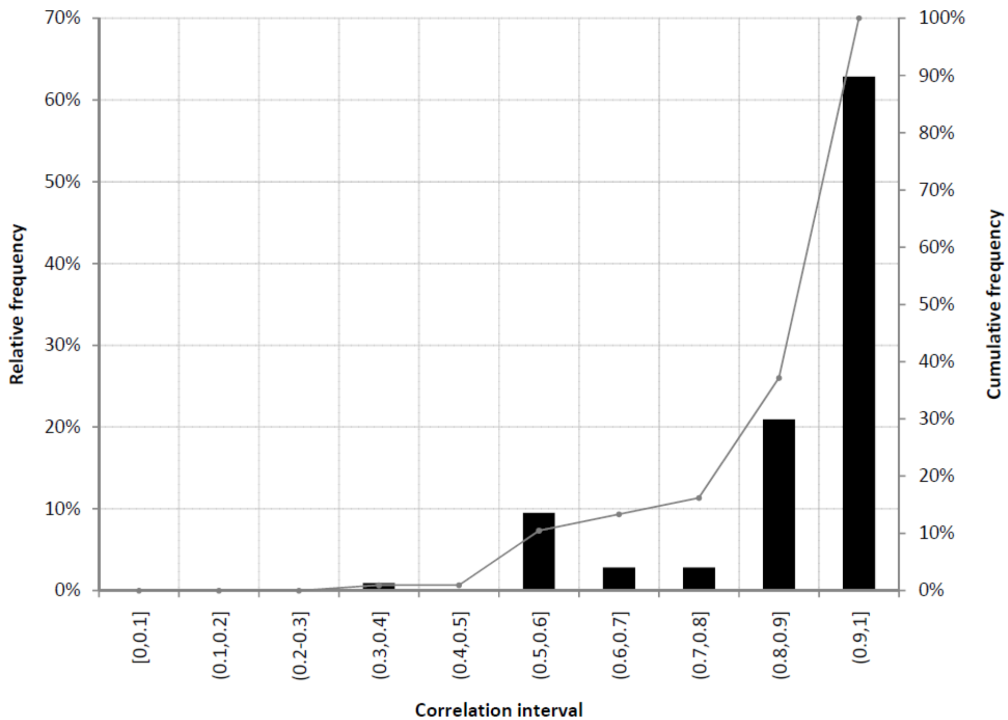
$$\rho^{\phi}_{X_1 X_2} = \frac{\sum_{i=1}^{n}(x_1(i) - \bar{x}_1)(x_2(i) - \bar{x}_2)}{\sqrt{\sum_{i=1}^{n}(x_1(i) - \bar{x}_1)^2 \sum_{i=1}^{n}(x_2(i) - \bar{x}_2)^2}}$$

- Relative risk:

$$\rho^{R}_{X_1 X_2} = \frac{(1 - \bar{x}_2)}{\bar{x}_2}\left(\frac{n\bar{x}_1}{\sum_{i=1}^{n} x_1(i)\,(1 - x_2(i))} - 1\right)$$

Correlated random vector generation: Motivation

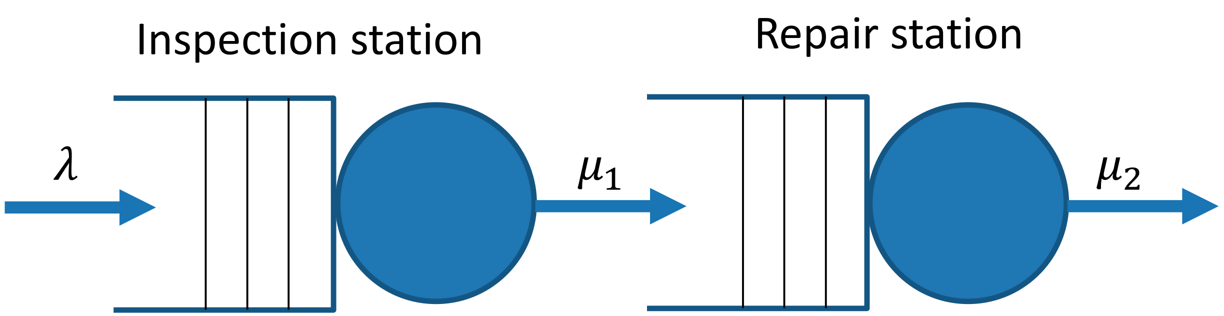
- We might want to capture the data properties, including correlation.
  - Portfolio optimization of oil & gas projects – semivariance (with J. Sefair and L. Zuluaga)



Universidad de los Andes

Correlated random vector generation: Motivation

- We might want to capture the data properties, including correlation.
  - Portfolio optimization of oil & gas projects – semivariance (with J. Sefair and L. Zuluaga)
  - Stochastic vehicle routing with correlated times (with J. Mendoza & R. Akhavan)
- Law & Kelton (2000):



  - Positive correlation of service times in tandem queues improves system performance (Mitchell et al.,1977)

Correlated random vector generation: Background

- How to generate correlated random vectors?
  - Generate random numbers with a priori correlation structure:

$$[x_1(1) \quad x_2(1) \quad \dots \quad x_k(1)]$$

Correlated random vector generation: Background

- How to generate correlated random vectors?
  - Generate random numbers with a priori correlation structure:

$$\begin{bmatrix} x_1(1) & x_2(1) & \cdots & x_k(1) \\ x_1(2) & x_2(2) & & x_k(2) \end{bmatrix}$$

Universidad de los Andes

Correlated random vector generation: Background

- How to generate correlated random vectors?
  - Generate random numbers with a priori correlation structure:

$$\begin{bmatrix} x_1(1) & x_2(1) & & x_k(1) \\ x_1(2) & x_2(2) & \cdots & x_k(2) \\ \vdots & \vdots & & \vdots \\ x_1(n) & x_2(k) & & x_k(n) \end{bmatrix}$$

  - A few remarks:
    - Prior knowledge of joint distributions
    - Take advantage of certain properties of specific distributions
    - What about a mixture of distributions?

Universidad de los Andes

Correlated random vector generation: Background

- How to generate correlated random vectors?
  - Generate random numbers with a priori correlation structure
  - Generate independent random numbers, **then** induce correlation structure

Correlated random vector generation: Background

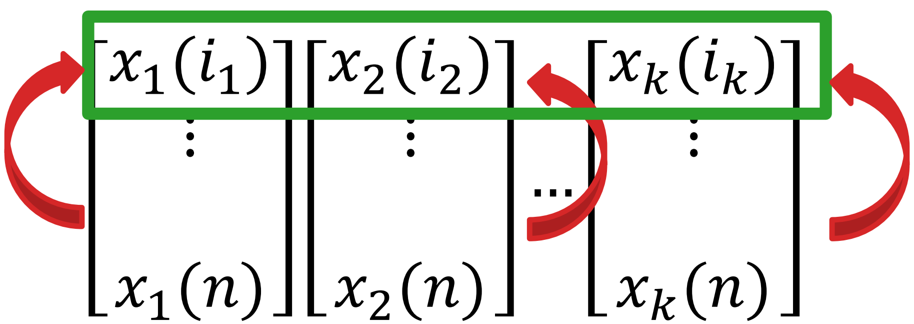
- How to generate correlated random vectors?
  - Generate random numbers with a priori correlation structure
  - Generate independent random numbers, **then** induce correlation structure

$$\begin{bmatrix} x_1(1) \\ x_1(2) \\ \vdots \\ x_1(n) \end{bmatrix} \begin{bmatrix} x_2(1) \\ x_2(2) \\ \vdots \\ x_2(n) \end{bmatrix} \cdots \begin{bmatrix} x_k(1) \\ x_k(2) \\ \vdots \\ x_k(n) \end{bmatrix}$$

$$f_{X_1}(x) \quad f_{X_2}(x) \quad f_{X_k}(x)$$

Correlated random vector generation: Background

- How to generate correlated random vectors?
  - Generate random numbers with a priori correlation structure
  - Generate independent random numbers, **then** induce correlation structure



- A few remarks:
  - @Risk and Crystal Ball induce correlation by Iman & Conover (1982) - (henceforth, IC)
  - Spearman rank correlation as a proxy of Pearson (linear) correlation.

Inducing bivariate correlation:
Example: inducing correlations in a data set (*n*=20)

$\bar{\rho}^{P}_{X_1X_2} = 0.8$

	X_1	X_2
1	1.2289	2.1680
2	6.4909	2.2104
3	0.2583	3.4869
4	1.3917	4.0804
5	0.8331	1.0123
6	2.1651	2.4693
7	1.5916	1.4274
8	2.3197	0.9236
9	5.1902	1.9912
10	3.2722	1.4160
11	1.2790	0.7229
12	1.3522	2.0748
13	3.0653	3.3528
14	0.8599	1.3222
15	3.0584	1.8016
16	0.9110	2.7710
17	1.7244	0.6049
18	1.4698	1.5123
19	0.8855	1.0728
20	4.5295	1.1795

$\rho^{P}_{X_1X_2} = -0.0197$

$\rho^{S}_{X_1X_2} = 0.0241$

	X_1	X_2
1	1.2289	1.0123
2	6.4909	4.0804
3	0.2583	1.1795
4	1.3917	0.9236
5	0.8331	1.9912
6	2.1651	1.8016
7	1.5916	0.7229
8	2.3197	3.3528
9	5.1902	3.4869
10	3.2722	2.4693
11	1.2790	0.6049
12	1.3522	1.3222
13	3.0653	2.1680
14	0.8599	2.0748
15	3.0584	2.7710
16	0.9110	1.5123
17	1.7244	1.4160
18	1.4698	1.4274
19	0.8855	1.0728
20	4.5295	2.2104

$\rho^{P}_{X_1X_2} = 0.7269$

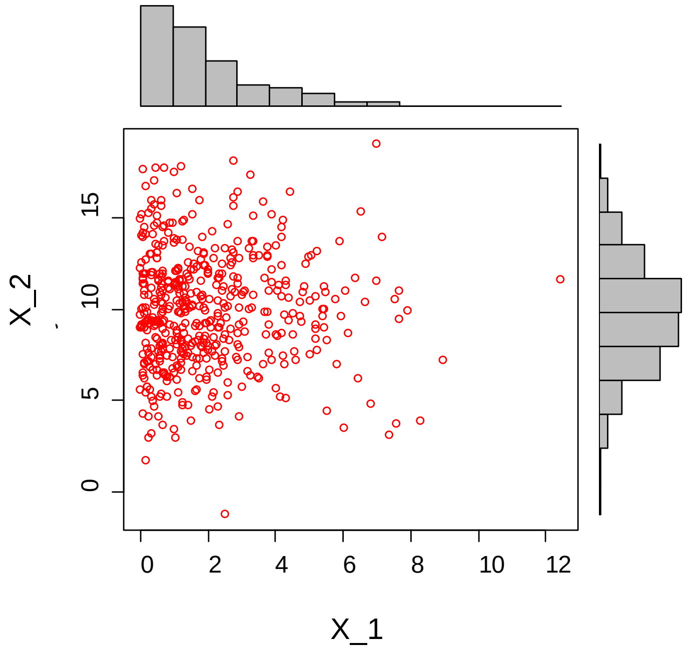
$\rho^{S}_{X_1X_2} = 0.8662$

	X_1	X_2
1	1.2289	1.0728
2	6.4909	3.3528
3	0.2583	0.6049
4	1.3917	1.3222
5	0.8331	1.1795
6	2.1651	2.0748
7	1.5916	3.4869
8	2.3197	1.4274
9	5.1902	4.0804
10	3.2722	2.1680
11	1.2790	1.4160
12	1.3522	1.5123
13	3.0653	2.2104
14	0.8599	0.9236
15	3.0584	2.7710
16	0.9110	1.0123
17	1.7244	1.8016
18	1.4698	1.9912
19	0.8855	0.7229
20	4.5295	2.4693

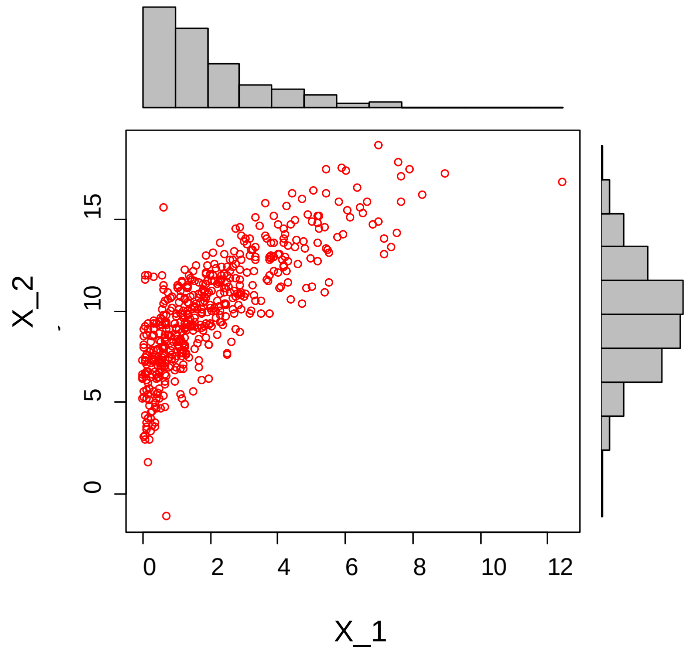
$\rho^{P}_{X_1X_2} = 0.8000$

$\rho^{S}_{X_1X_2} = 0.8887$

Inducing bivariate correlation:
Example: inducing correlations in a data set (n=500)



$$\rho^{P}_{X_1 X_2} = 0.0$$



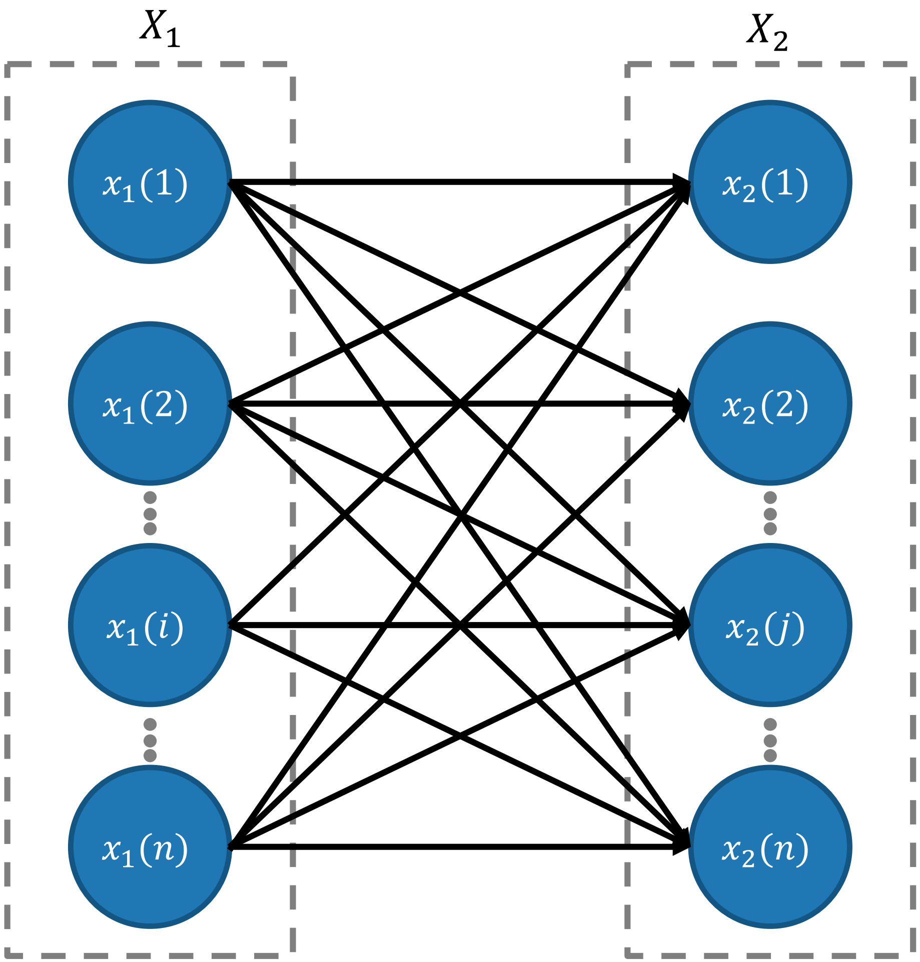
$$\rho^{P}_{X_1 X_2} = 0.8$$

Note: These are the same 500 observations of an exponential and a normal distribution

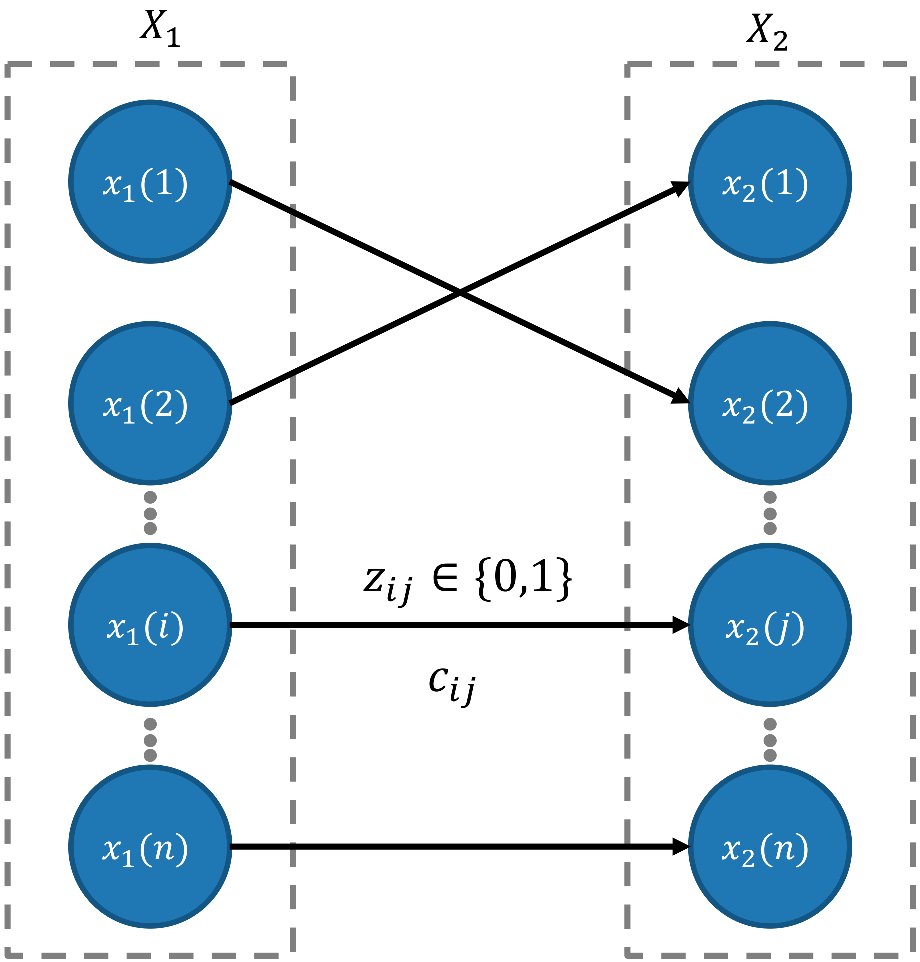
Outline

- Correlated random vector generation
- MIP model
- Column generation
- Concluding remarks
- Future work
- References

MIP model: Intuition



MIP model: Intuition



Universidad de los Andes

MIP model: Cost definition

- Spearman rank correlation:

$$c_{ij} = \left(r_{X_1}(x_1(i)) - r_{X_2}(x_2(j))\right)^2$$

- Pearson correlation:

$$c_{ij} = (x_1(i) - \bar{x}_1)(x_2(j) - \bar{x}_2)$$

- Kendall's coefficient of concordance

$$c_{ij} = (x_1(i) + x_2(j))^2$$

- Phi coefficient of correlation (for categorical variables):

$$c_{ij} = (x_1(i) - \bar{x}_1)(x_2(j) - \bar{x}_2)$$

- Relative risk:

$$c_{ij} = x_1(i)\,(1 - x_2(j))$$

MIP model: Formulation

$$\min \delta^+ + \delta^-$$

Minimize deviation from target

s. t.

$$\frac{\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij} z_{ij}}{\sqrt{\sum_{i=1}^{n}(x_1(i)-\bar{x}_1)^2 \sum_{i=1}^{n}(x_2(i)-\bar{x}_2)^2}} + \delta^- - \delta^+ = \bar{\rho}^P_{X_1 X_2}$$

Sample correlation

Target correlation

$$\sum_{j=1}^{n} z_{ij} = 1, \qquad i = 1, \dots, n$$

$$\sum_{i=1}^{n} z_{ij} = 1, \qquad j = 1, \dots, n$$

$$z_{ij} \in \{0,1\}, \qquad i, j = 1, \dots, n$$

Assignment problem

$$\delta^+, \delta^- \geq 0$$

MIP model: Computational experiments

- Samples per combination ($\bar{\rho}^P$, n, distributions): 10
  - $\bar{\rho}^P$=-0.8, -0.5, 0, 0.5, 0.8
  - Sample sizes (n) of 500, 1000, 2000, 3000
  - Combinations of exponential, normal and uniform **distributions**: (exp.-exp., normal-normal, unif.-unif., exp.-normal, exp.-unif., normal-unif.)
- Coded in Java (v. 1.8.0_45) with Gurobi (v. 6.0.5)
  - Time limit of 1000 seconds
- Intel Xeon double-core processor @2.40GHz and 32GB of RAM

MIP model:
Computational results: exp.-exp. & n=1000

$\bar{\rho}^P$	Presolve (s)	LR (s)	B&B (s)	Total (s)	LR(%)	$\delta^+ + \delta^-$	MIP ρ^P	IC ρ^S	IC ρ^P
-0.8	15.00	86.72	2.40	104.12	83.3%	0.153	-0.647	-0.815	-0.552
-0.5	14.92	26.61	2.00	43.53	61.1%	0.000	-0.500	-0.436	-0.352
0	15.13	2.03	9.40	26.56	7.6%	0.000	0.000	-0.051	-0.042
0.5	16.11	20.28	13.25	49.64	40.9%	0.000	0.500	0.490	0.511
0.8	15.22	30.33	16.25	61.80	49.1%	0.000	0.800	0.797	0.699

Gurobi (v. 6.0.5)

MIP model:
Computational results: exp.-exp. & n=1000

MIP solution is closer to the target than IC, since IC uses Spearman Rank correlation

$\bar{\rho}^P$	Presolve (s)	LR (s)	B&B (s)	Total (s)	LR(%)	$\delta^+ + \delta^-$	MIP ρ^P	IC ρ^S	IC ρ^P
-0.8	15.00	86.72	2.40	104.12	83.3%	0.153	-0.647	-0.815	-0.552
-0.5	14.92	26.61	2.00	43.53	61.1%	0.000	-0.500	-0.436	-0.352
0	15.13	2.03	9.40	26.56	7.6%	0.000	0.000	-0.051	-0.042
0.5	16.11	20.28	13.25	49.64	40.9%	0.000	0.500	0.490	0.511
0.8	15.22	30.33	16.25	61.80	49.1%	0.000	0.800	0.797	0.699

Gurobi (v. 6.0.5)

Universidad de los Andes

MIP model: Computational results: exp.-exp. & n=1000

MIP solution is closer to the target than IC, since IC uses Spearman Rank correlation

$\bar{\rho}^P$	Presolve (s)	LR (s)	B&B (s)	Total (s)	LR(%)	$\delta^+ + \delta^-$	MIP ρ^P	IC ρ^S	IC ρ^P
-0.8	15.00	86.72	2.40	104.12	83.3%	0.153	-0.647	-0.815	-0.552
-0.5	14.92	26.61	2.00	43.53	61.1%	0.000	-0.500	-0.436	-0.352
0	15.13	2.03	9.40	26.56	7.6%	0.000	0.000	-0.051	-0.042
0.5	16.11	20.28	13.25	49.64	40.9%	0.000	0.500	0.490	0.511
0.8	15.22	30.33	16.25	61.80	49.1%	0.000	0.800	0.797	0.699

Gurobi (v. 6.0.5)

Universidad de los Andes

MIP model: Computational results: exp.-exp. & n=1000

This case is especially challenging, as it is impossible to achieve (Conway, 1979)

$\bar{\rho}^P$	Presolve (s)	LR (s)	B&B (s)	Total (s)	LR(%)	$\delta^+ + \delta^-$	MIP ρ^P	IC ρ^S	IC ρ^P
-0.8	15.00	86.72	2.40	104.12	83.3%	0.153	-0.647	-0.815	-0.552
-0.5	14.92	26.61	2.00	43.53	61.1%	0.000	-0.500	-0.436	-0.352
0	15.13	2.03	9.40	26.56	7.6%	0.000	0.000	-0.051	-0.042
0.5	16.11	20.28	13.25	49.64	40.9%	0.000	0.500	0.490	0.511
0.8	15.22	30.33	16.25	61.80	49.1%	0.000	0.800	0.797	0.699

Gurobi (v. 6.0.5)

MIP model: Computational results: exp.-exp. & n=1000

Room for improvement: Simplex iterations

$\bar{\rho}^P$	Presolve (s)	LR (s)	B&B (s)	Total (s)	LR(%)	$\delta^+ + \delta^-$	MIP ρ^P	IC ρ^S	IC ρ^P
-0.8	15.00	86.72	2.40	104.12	83.3%	0.153	-0.647	-0.815	-0.552
-0.5	14.92	26.61	2.00	43.53	61.1%	0.000	-0.500	-0.436	-0.352
0	15.13	2.03	9.40	26.56	7.6%	0.000	0.000	-0.051	-0.042
0.5	16.11	20.28	13.25	49.64	40.9%	0.000	0.500	0.490	0.511
0.8	15.22	30.33	16.25	61.80	49.1%	0.000	0.800	0.797	0.699

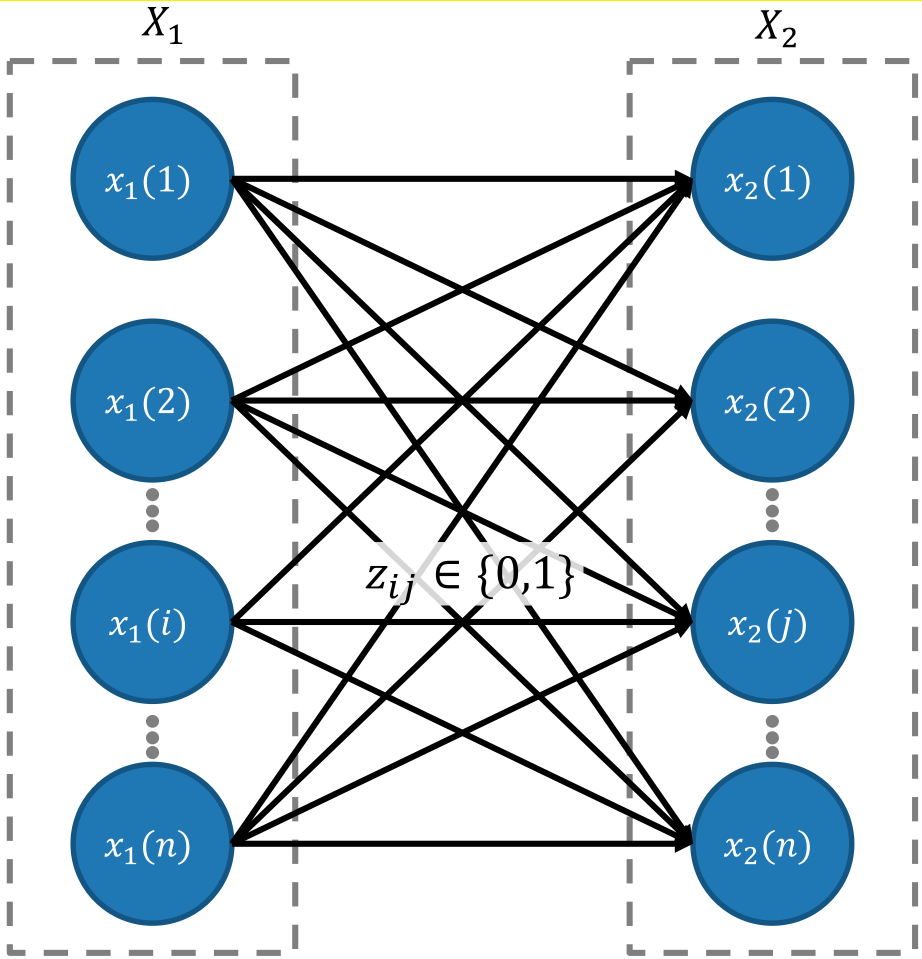
Gurobi (v. 6.0.5)

- How to speed up the linear relaxation?

Outline

- Correlated random vector generation
- MIP model
- Column generation
- Concluding remarks
- Future work
- References

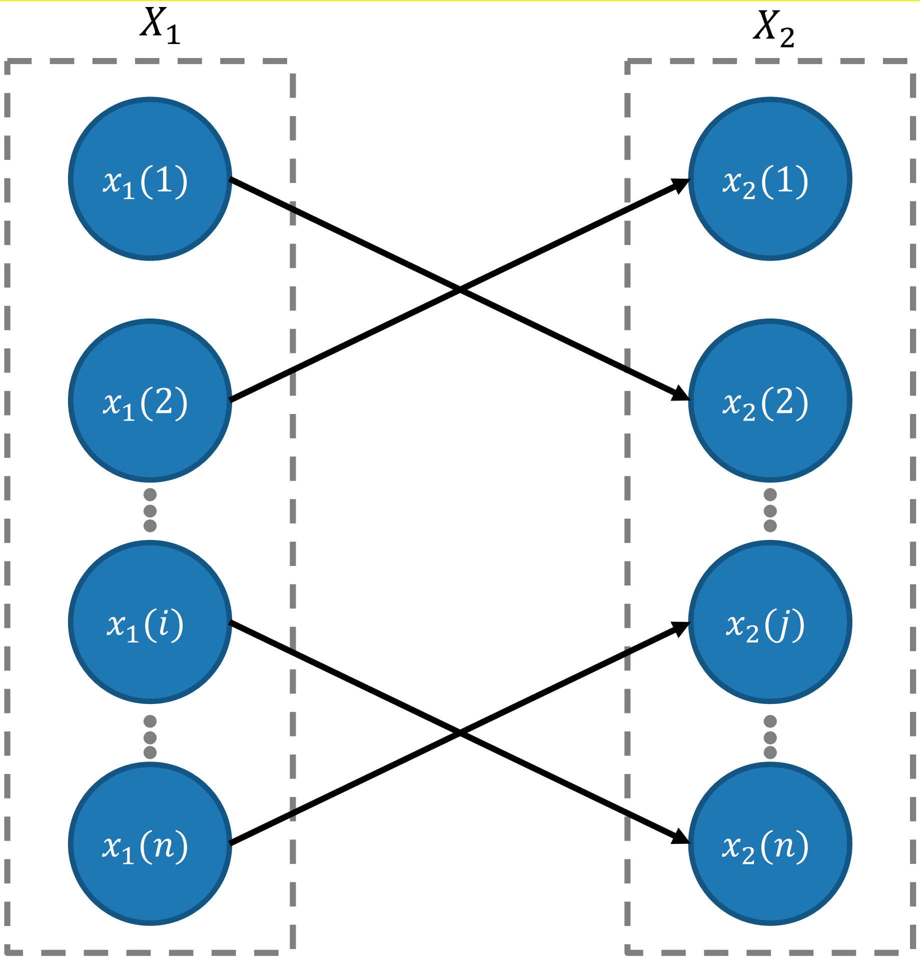
Column generation
Intuition



n^2 total z_{ij} variables, yet only n of them are nonzero in the optimal solution

Column generation

Intuition



n^2 total z_{ij} variables, yet only n of them are nonzero in the optimal solution

Column generation: Original problem

$$\min \delta^+ + \delta^-$$

s. t.

$$\frac{\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij} z_{ij}}{\sqrt{\sum_{i=1}^{n}(x_1(i)-\bar{x}_1)^2 \sum_{i=1}^{n}(x_2(i)-\bar{x}_2)^2}} + \delta^- - \delta^+ = \bar{\rho}^P_{X_1X_2}$$

$$\sum_{j=1}^{n} z_{ij} = 1, \qquad i = 1, \dots, n$$

$$S_{X_1X_2} = \sqrt{\sum_{i=1}^{n}(x_1(i)-\bar{x}_1)^2 \sum_{i=1}^{n}(x_2(i)-\bar{x}_2)^2}$$

$$\sum_{i=1} \qquad j = 1, \dots, n$$

$$z_{ij} \in \{0,1\}, \qquad i,j = 1, \dots, n$$

$$\delta^+, \delta^- \geq 0$$

Universidad de los Andes

Column generation: Original problem

$$\min \delta^+ + \delta^-$$

s.t.

$$\frac{\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij} z_{ij}}{S_{X_1 X_2}} + \delta^- - \delta^+ = \bar{\rho}^P_{X_1 X_2}$$

$$\sum_{j=1}^{n} z_{ij} = 1, \qquad i = 1, \dots, n$$

$$\sum_{i=1}^{n} z_{ij} = 1, \qquad j = 1, \dots, n$$

$$z_{ij} \in \{0,1\}, \qquad i, j = 1, \dots, n$$

$$\delta^+, \delta^- \geq 0$$

Column generation:
Relaxation and column structure

$$\min \delta^+ + \delta^-$$

s. t.

$$\frac{\sum_{i=1}^n \sum_{j=1}^n c_{ij} z_{ij}}{S_{X_1 X_2}} + \delta^- - \delta^+ = \bar{\rho}^P_{X_1 X_2}$$

$$\sum_{j=1}^n z_{ij} = 1, \qquad i = 1, \dots, n$$

$$\sum_{i=1}^n z_{ij} = 1, \qquad j = 1, \dots, n$$

$$z_{ij} \geq 0, \qquad i, j = 1, \dots, n$$

$$\delta^+, \delta^- \geq 0$$

A new z_{ij}:

$$\begin{bmatrix} c_{ij}/S_{X_1 X_2} \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix}$$

Universidad de los Andes

Column generation: Dual variables

$$\min \delta^+ + \delta^-$$

s.t.

$$\frac{\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij} z_{ij}}{S_{X_1 X_2}} + \delta^- - \delta^+ = \bar{\rho}^P_{X_1 X_2} \qquad [\sigma]$$

$$\sum_{j=1}^{n} z_{ij} = 1, \qquad i = 1, \dots, n \qquad [u_i]$$

$$\sum_{i=1}^{n} z_{ij} = 1, \qquad j = 1, \dots, n \qquad [v_j]$$

$$z_{ij} \geq 0, \qquad i, j = 1, \dots, n$$

$$\delta^+, \delta^- \geq 0$$

Column generation: Reduced costs

$$r_q = c_q - \mathbf{w}^T \mathbf{a}_q, \qquad q \in I_N$$

$$r_{(i,j)} = 0 - [\sigma \quad \cdots \quad u_i \quad \cdots \quad v_j \quad \cdots] \begin{bmatrix} c_{ij}/S_{X_1X_2} \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{bmatrix}$$

$$r_{(i,j)} = -\frac{c_{ij}}{S_{X_1X_2}}\sigma - u_i - v_j$$

Column generation:
Choosing entering columns

$$\min_{i,j\in\{1,\ldots,n\}}\left\{-\frac{c_{ij}}{S_{X_1X_2}}\sigma-u_i-v_j\right\}$$

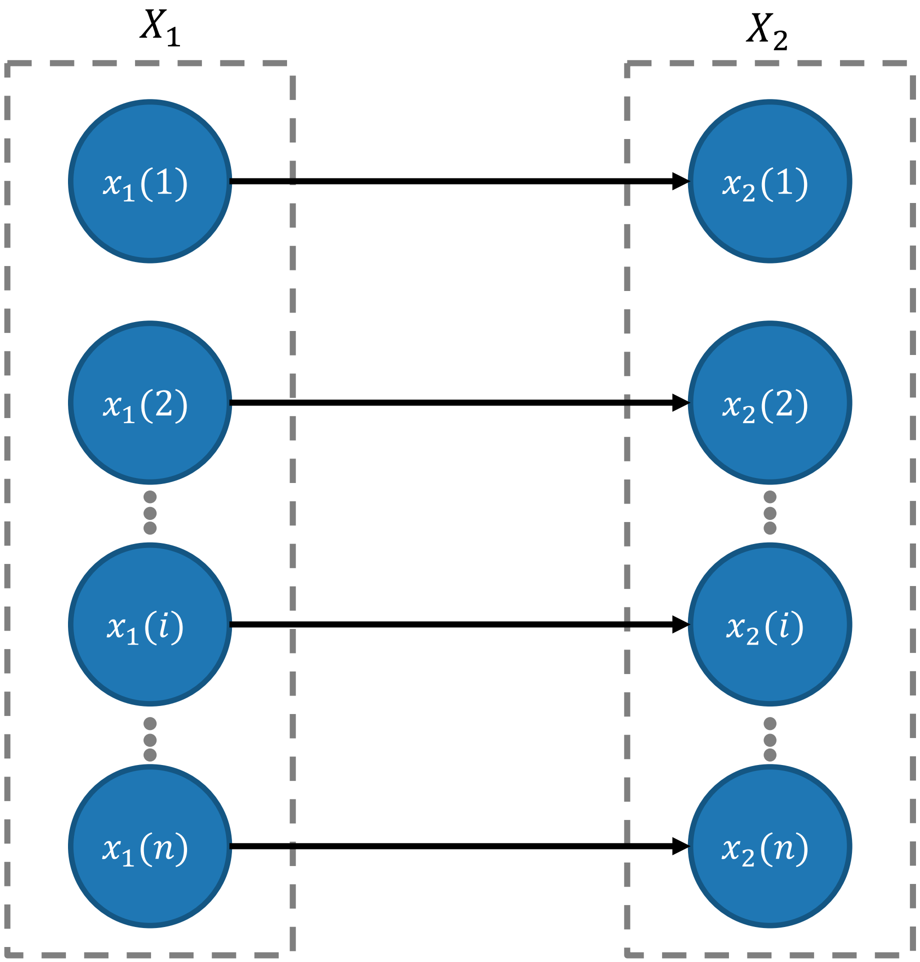
Closed form, easy $r_{(i,j)}$ calculation

Universidad de los Andes

Column generation: Implementation

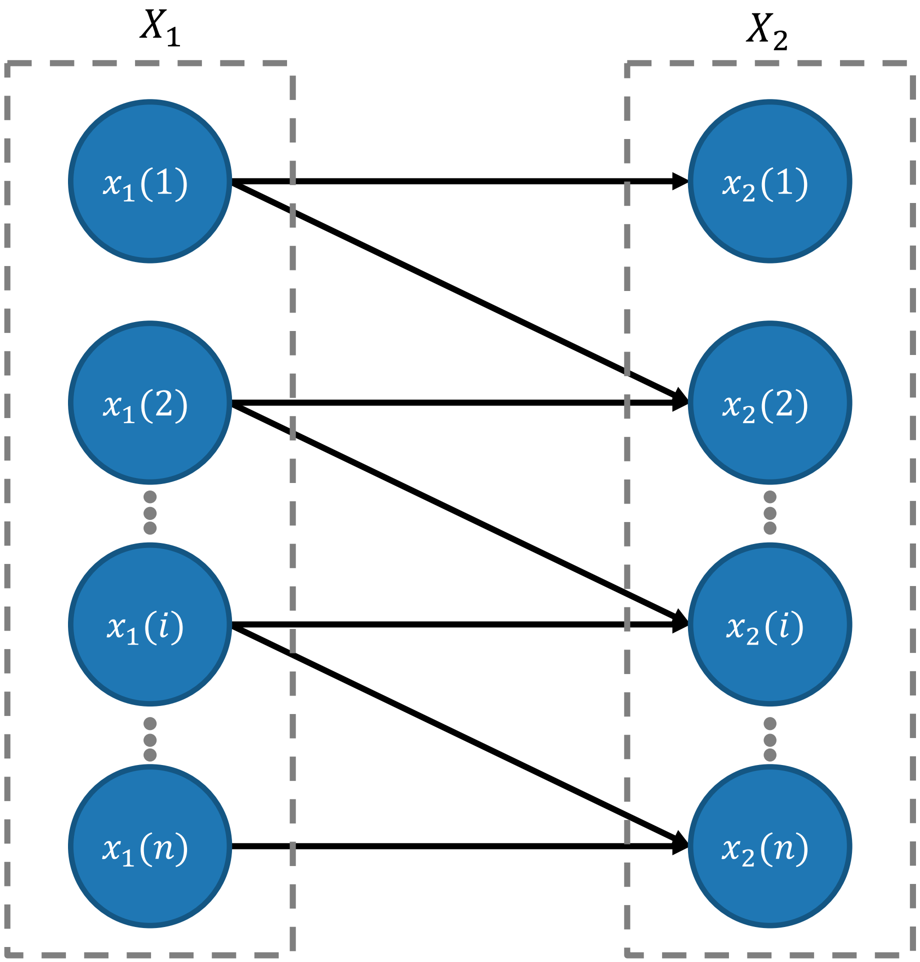
- At every iteration:
  - compute reduced costs for every nonbasic variable z_{ij}
  - add z_{ij} with the smallest reduced cost
- At the end of column generation:
  - Force integrality and solve MIP

Column generation: Initialization



$$z_{ii}, i \in \{1, \dots, n\}$$

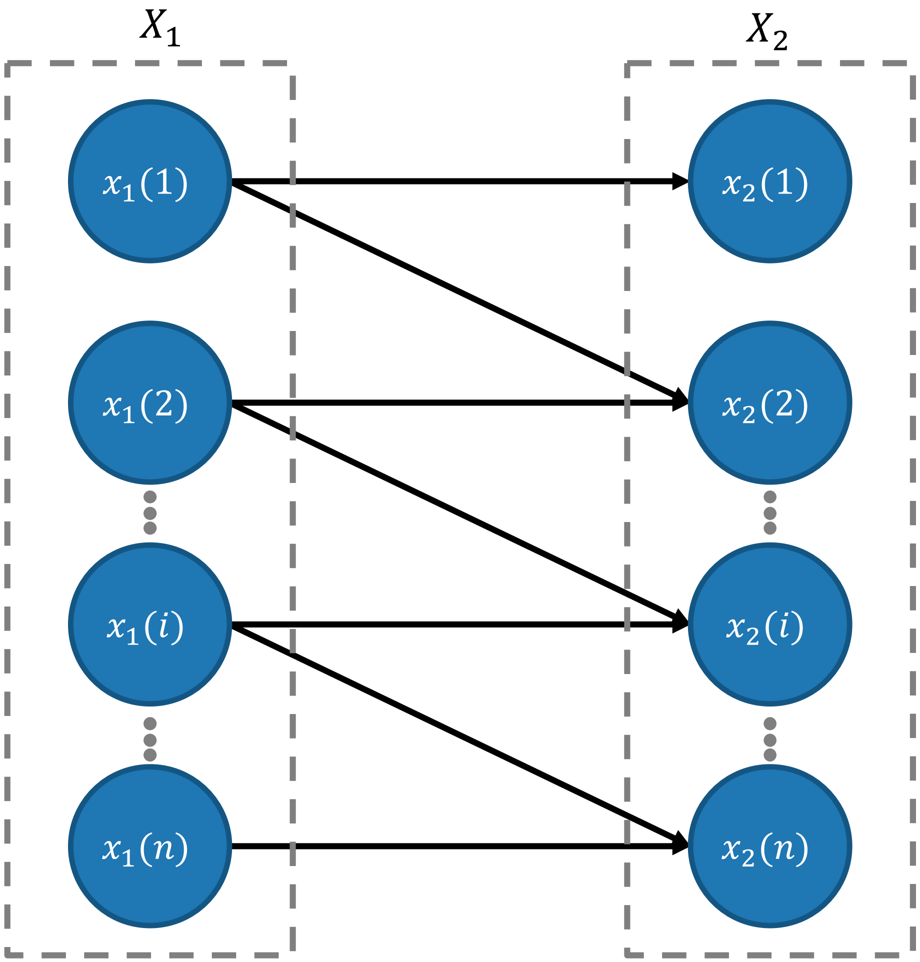
Column generation: Initialization



$$z_{ii}, i \in \{1, \dots, n\}$$

$$z_{i,i+1}, i \in \{1, \dots, n-1\}$$

Column generation: Initialization



$$z_{ii}, i \in \{1, \dots, n\}$$
$$z_{i,i+1}, i \in \{1, \dots, n-1\}$$
$$\delta^+, \delta^-$$

Universidad de los Andes

Column generation: A note on initialization

- Initial variables depend on the order of the data set
- What if we use the IC output as the initial sorting?
- Random vs. sorted input (for n=1000, normal – uniform)

$\bar{\rho}^P$	Avg. number of generated columns		$\delta^+ + \delta^-$		Computational time (s)	
	Random Sorting	IC sorting	Random Sorting	IC sorting	Random Sorting	IC sorting
-0.8	778.2	232.2	0.000	0.000	14.06	4.90
-0.5	327.4	2.0	0.011	0.000	5.57	0.11
0	2.8	2.3	0.005	0.013	0.10	0.13
0.5	314.9	2.0	0.010	0.000	5.35	0.12
0.8	773.0	193.7	0.000	0.000	14.64	4.11

Computational results: Average number of generated columns

Sample Size	$\overline{\rho}$	Distributions					
		exp-exp	norm-norm	unif-unif	exp-norm	exp-unif	norm-unif
500	-0.8		307.4	115.4	358.1	414.5	257.2
	-0.5	244.8	66.7	102.6	141.5	135.6	40.2
	0	3.8	2.3	2.1	2.7	3.0	2.6
	0.5	2.0	2.0	2.0	2.0	48.8	2.0
	0.8	270.2	118.6	2.0	326.3	410.2	13.0
1000	-0.8		564.1	2.0	674.7	827.3	232.2
	-0.5	417.1	3.9	71.2	179.8	272	2.0
	0	2.2	2.8	2.3	3.1	3.1	2.3
	0.5	122.9	3.6	80.8	154.2	260.8	2.0
	0.8	519.6	474.3	2.0	678.1	838.3	193.7
2000	-0.8		1222.0	27.9	1361	1699.9	640.1
	-0.5	826.8	13.6	132.6	324.9	518	2.0
	0	2.6	2.1	2.6	2.4	2.8	3.0
	0.5	302.7	2.0	153.3	424.0	577.4	2.0
	0.8	985.1	1036.4	2.0	1433.8	1710.3	450.5
3000	-0.8		1826.4	22.2	2186.9	2594.1	1033.9
	-0.5	1295.1	84.6	340.9	650	888.1	7.3
	0	3.3	2.0	2.1	2.2	2.2	2.1
	0.5	319.2	2.0	110.8	459.8	727.3	2.0
	0.8	1486.5	1577.1	2.0	2131.1	2536.3	490.4

Computational results: Average number of generated columns

Sample Size	$\bar{\rho}$	Distributions					
		exp-exp	norm-norm	unif-unif	exp-norm	exp-unif	norm-unif
500	-0.8		307.4	115.4	358.1	414.5	257.2
500	-0.5	244.8	66.7	102.6	141.5	135.6	40.2
500	0	3.8	2.3	2.1	2.7	3.0	2.6
500	0.5	2.0	2.0	2.0	2.0	48.8	2.0
500	0.8	270.2	118.6	2.0	326.3	410.2	13.0
1000	-0.8		564.1	2.0	674.7	827.3	232.2
1000	-0.5	417.1	3.9	71.2	179.8	272	2.0
1000	0	[illegible]	[illegible]	[illegible]	3.1	3.1	2.3
1000	0.5	[illegible]	[illegible]	[illegible]	154.2	260.8	2.0
1000	0.8	[illegible]	[illegible]	[illegible]	678.1	838.3	193.7
2000	-0.8	[illegible]	[illegible]	[illegible]	1361	1699.9	640.1
2000	-0.5	[illegible]	[illegible]	[illegible]	324.9	518	2.0
2000	0	2.6	2.1	2.6	2.4	2.8	3.0
2000	0.5	302.7	2.0	153.3	424.0	577.4	2.0
2000	0.8	985.1	1036.4	2.0	1433.8	1710.3	450.5
3000	-0.8		1826.4	22.2	2186.9	2594.1	1033.9
3000	-0.5	1295.1	84.6	340.9	650	888.1	7.3
3000	0	3.3	2.0	2.1	2.2	2.2	2.1
3000	0.5	319.2	2.0	110.8	459.8	727.3	2.0
3000	0.8	1486.5	1577.1	2.0	2131.1	2536.3	490.4

For all cases, the average number of generated columns is $< n$

Universidad de los Andes

Computational results: Average fraction of generated columns / sample size

Sample Size	$\overline{\rho}$	Distributions					
		exp-exp	norm-norm	unif-unif	exp-norm	exp-unif	norm-unif
500	-0.8		0.615	0.231	0.716	0.829	0.514
	-0.5	0.490	0.133	0.205	0.283	0.271	0.080
	0	0.008	0.005	0.004	0.005	0.006	0.005
	0.5	0.004	0.004	0.004	0.004	0.098	0.004
	0.8	0.540	0.237	0.004	0.653	0.820	0.026
1000	-0.8		0.564	0.002	0.675	0.827	0.232
	-0.5	0.417	0.004	0.071	0.180	0.272	0.002
	0	0.002	0.003	0.002	0.003	0.003	0.002
	0.5	0.123	0.004	0.081	0.154	0.261	0.002
	0.8	0.520	0.474	0.002	0.678	0.838	0.194
2000	-0.8		0.611	0.014	0.681	0.850	0.320
	-0.5	0.413	0.007	0.066	0.162	0.259	0.001
	0	0.001	0.001	0.001	0.001	0.001	0.002
	0.5	0.151	0.001	0.077	0.212	0.289	0.001
	0.8	0.493	0.518	0.001	0.717	0.855	0.225
3000	-0.8		0.609	0.007	0.729	0.865	0.345
	-0.5	0.432	0.028	0.114	0.217	0.296	0.002
	0	0.001	0.001	0.001	0.001	0.001	0.001
	0.5	0.106	0.001	0.037	0.153	0.242	0.001
	0.8	0.496	0.526	0.001	0.710	0.845	0.163

Computational results: Deviation from target correlation

Sample size	$\bar{\rho}$	exp-exp		norm-norm		unif-unif		exp-norm		exp-unif		norm-unif	
		MIP	CG	MIP	CG	MIP	CG	MIP	CG	MIP	CG	MIP	CG
500	-0.8			0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	-0.5	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.012	0.000	0.002	0.000	0.000
	0	0.001	0.043	0.001	0.032	0.002	0.043	0.001	0.035	0.002	0.045	0.001	0.041
	0.5	0.002	0.052	0.001	0.039	0.001	0.014	0.001	0.021	0.000	0.000	0.002	0.046
	0.8	0.000	0.000	0.000	0.000	0.002	0.013	0.000	0.000	0.000	0.000	0.001	0.002
1000	-0.8			0.000	0.000	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000
	-0.5	0.000	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.011
	0	0.000	0.002	0.000	0.005	0.000	0.002	0.000	0.006	0.000	0.008	0.001	0.005
	0.5	0.000	0.000	0.001	0.005	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.010
	0.8	0.000	0.000	0.000	0.000	0.001	0.003	0.000	0.000	0.000	0.000	0.000	0.000
2000	-0.8			0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	-0.5	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.010
	0	0.000	0.002	0.000	0.016	0.000	0.007	0.000	0.004	0.000	0.003	0.001	0.011
	0.5	0.000	0.000	0.000	0.008	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.014
	0.8	0.000	0.000	0.000	0.000	0.000	0.004	0.000	0.000	0.000	0.000	0.000	0.000
3000	-0.8			0.000	0.000	0.000	0.000	-	0.000	0.000	0.000	-	0.000
	-0.5	0.000	0.000	0.000	0.000	0.000	0.000	-	0.000	0.000	0.001	-	0.001
	0	0.000	0.005	0.000	0.009	0.000	0.016	-	0.006	0.000	0.012	-	0.011
	0.5	0.000	0.000	0.001	0.014	0.000	0.000	-	0.000	0.000	0.000	-	0.024
	0.8	0.000	0.000	0.000	0.000	0.000	0.005	-	0.000	0.000	0.000	-	0.000

Computational results: Deviation from target correlation

Sample size	$\overline{\rho}$	exp-exp MIP	exp-exp CG	norm-norm MIP	norm-norm CG	unif-unif MIP	unif-unif CG	exp-norm MIP	exp-norm CG	exp-unif MIP	exp-unif CG	norm-unif MIP	norm-unif CG
500	-0.8			0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	-0.5	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.012	0.000	0.002	0.000	0.000
	0	0.001	0.043	0.001	0.032	0.002	0.043	0.001	0.035	0.002	0.045	0.001	0.041
	0.5	0.002	0.052	0.001	0.039	0.001	0.014	0.001	0.021	0.000	0.000	0.002	0.046
	0.8	0.000	0.000	0.000	0.000	0.002	0.013	0.000	0.000	0.000	0.000	0.001	0.002
1000	-0.8			0.000	0.000	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000
	-0.5	0.000	0.000	0.000	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.011
	0	0.000	0.002	0.000	0.005	0.000	0.002	0.000	0.006	0.000	0.008	0.001	0.005
	0.5	0.000	0.000	0.001	0.005	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.010
	0.8	0.000	0.000	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.000	0.000	0.000
2000	-0.8			[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.000	0.000	0.000
	-0.5	0.000	0.000	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.000	0.001	0.010
	0	0.000	0.002	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.003	0.001	0.011
	0.5	0.000	0.000	0.000	0.008	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.014
	0.8	0.000	0.000	0.000	0.000	0.000	0.004	0.000	0.000	0.000	0.000	0.000	0.000
3000	-0.8			0.000	0.000	0.000	0.000	-	0.000	0.000	0.000	-	0.000
	-0.5	0.000	0.000	0.000	0.000	0.000	0.000	-	0.000	0.000	0.001	-	0.001
	0	0.000	0.005	0.000	0.009	0.000	0.016	-	0.006	0.000	0.012	-	0.011
	0.5	0.000	0.000	0.001	0.014	0.000	0.000	-	0.000	0.000	0.000	-	0.024
	0.8	0.000	0.000	0.000	0.000	0.000	0.005	-	0.000	0.000	0.000	-	0.000

Loss of quality is minimal ($\sim \mathbf{1} \times \mathbf{10^{-2}}$)

Computational results: Average time (in seconds)

Sample size	$\bar{\rho}$	exp-exp		norm-norm		unif-unif		exp-norm		exp-unif		norm-unif	
		MIP	CG	MIP	CG	MIP	CG	MIP	CG	MIP	CG	MIP	CG
500	-0.8			13.57	2.05	11.78	0.83	14.60	1.95	15.93	2.61	13.38	1.62
	-0.5	7.07	1.62	9.86	0.44	10.67	0.76	9.53	0.90	10.38	0.86	10.01	0.28
	0	4.49	0.05	6.67	0.03	11.00	0.04	7.86	0.04	6.20	0.05	8.94	0.04
	0.5	5.26	0.04	9.17	0.03	9.38	0.03	8.90	0.04	9.71	0.31	9.75	0.04
	0.8	6.82	1.87	12.53	0.75	12.45	0.04	16.69	1.92	17.12	2.37	12.51	0.09
1000	-0.8			136.59	11.85	154.33	0.09	170.15	14.11	178.55	17.29	153.93	4.90
	-0.5	70.87	8.43	88.06	0.16	89.83	1.56	98.03	3.68	90.88	5.90	109.32	0.11
	0	24.95	0.11	47.90	0.13	49.85	0.13	48.93	0.13	46.63	0.14	54.58	0.13
	0.5	40.40	2.54	89.93	0.14	91.03	1.79	104.32	3.23	103.06	5.50	110.17	0.12
	0.8	75.63	10.32	145.81	9.59	124.89	0.09	179.28	14.02	190.83	18.07	154.32	4.11
2000	-0.8			562.33	83.22	713.07	2.21	675.02	83.89	802.34	107.46	668.62	39.62
	-0.5	537.87	52.51	354.43	0.92	411.69	9.52	349.42	20.90	380.40	32.63	404.52	0.33
	0	142.83	0.37	135.59	0.33	173.38	0.42	134.22	0.34	121.30	0.41	157.40	0.37
	0.5	297.60	19.15	348.94	0.31	447.80	10.87	346.68	26.94	374.64	36.70	413.48	0.29
	0.8	516.91	61.21	556.81	71.08	674.69	0.34	690.57	87.15	813.46	107.88	675.99	27.65
3000	-0.8			1001.00	231.36	1000.90	3.16	1001.15	281.12	1001.30	333.10	1000.23	135.02
	-0.5	1001.29	174.62	1001.52	10.65	1001.05	44.19	1001.36	84.39	1001.22	113.19	1001.92	1.30
	0	725.43	0.78	487.79	0.62	418.91	0.72	502.18	0.63	621.75	0.64	724.31	0.66
	0.5	1001.51	42.51	1001.14	0.65	1001.45	14.45	1001.15	59.71	1001.76	91.87	1000.37	0.63
	0.8	1001.38	198.90	1001.04	198.79	1000.96	0.62	1001.67	273.49	1001.41	325.90	1001.18	63.50

Computational results: Average time (in seconds)

Sample size	$\bar{\rho}$	exp-exp MIP	exp-exp CG	norm-norm MIP	norm-norm CG	unif-unif MIP	unif-unif CG	exp-norm MIP	exp-norm CG	exp-unif MIP	exp-unif CG	norm-unif MIP	norm-unif CG
500	-0.8			13.57	2.05	11.78	0.83	14.60	1.95	15.93	2.61	13.38	1.62
	-0.5	7.07	1.62	9.86	0.44	10.67	0.76	9.53	0.90	10.38	0.86	10.01	0.28
	0	4.49	0.05	6.67	0.03	11.00	0.04	7.86	0.04	6.20	0.05	8.94	0.04
	0.5	5.26	0.04	9.17	0.03	9.38	0.03	8.90	0.04	9.71	0.31	9.75	0.04
	0.8	6.82	1.87	12.53	0.75	12.45	0.04	16.69	1.92	17.12	2.37	12.51	0.09
1000	-0.8			136.59	11.85	154.33	0.09	170.15	14.11	178.55	17.29	153.93	4.90
	-0.5	70.87	8.43	88.06	0.16	89.83	1.56	98.03	3.68	90.88	5.90	109.32	0.11
	0	24.95	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	0.13	46.63	0.14	54.58	0.13
	0.5	40.40	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	3.23	103.06	5.50	110.17	0.12
	0.8	75.63	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	14.02	190.83	18.07	154.32	4.11
2000	-0.8			[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	83.89	802.34	107.46	668.62	39.62
	-0.5	537.87	52.51	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	20.90	380.40	32.63	404.52	0.33
	0	142.83	0.37	135.59	0.33	173.38	0.42	134.22	0.34	121.30	0.41	157.40	0.37
	0.5	297.60	19.15	348.94	0.31	447.80	10.87	346.68	26.94	374.64	36.70	413.48	0.29
	0.8	516.91	61.21	556.81	71.08	674.69	0.34	690.57	87.15	813.46	107.88	675.99	27.65
3000	-0.8			1001.00	231.36	1000.90	3.16	1001.15	281.12	1001.30	333.10	1000.23	135.02
	-0.5	1001.29	174.62	1001.52	10.65	1001.05	44.19	1001.36	84.39	1001.22	113.19	1001.92	1.30
	0	725.43	0.78	487.79	0.62	418.91	0.72	502.18	0.63	621.75	0.64	724.31	0.66
	0.5	1001.51	42.51	1001.14	0.65	1001.45	14.45	1001.15	59.71	1001.76	91.87	1000.37	0.63
	0.8	1001.38	198.90	1001.04	198.79	1000.96	0.62	1001.67	273.49	1001.41	325.90	1001.18	63.50

Significant speedup vs MIP

Universidad de los Andes

Computational results: Average speedup vs MIP

Sample size	$\overline{\rho}$	exp-exp	norm-norm	unif-unif	exp-norm	exp-unif	norm-unif
500	-0.8		6.70	32.14	7.60	6.46	8.81
	-0.5	4.60	26.60	14.61	11.19	12.63	39.65
	0	101.25	207.27	289.81	209.41	167.79	231.61
	0.5	154.75	304.55	285.50	268.00	47.41	263.91
	0.8	3.64	25.39	377.43	8.63	7.50	290.64
1000	-0.8		11.76	1727.82	12.23	10.36	35.16
	-0.5	8.58	712.51	60.11	29.85	16.51	1012.82
	0	240.65	419.00	399.68	389.88	378.11	447.11
	0.5	17.97	825.82	57.70	38.87	19.22	980.36
	0.8	7.39	15.78	1361.84	13.05	10.73	52.56
2000	-0.8		6.80	1220.16	8.17	7.58	17.61
	-0.5	10.57	695.65	44.07	17.50	11.85	1243.75
	0	402.83	431.52	421.11	415.26	312.00	446.82
	0.5	15.77	1138.34	44.57	13.05	10.34	1484.24
	0.8	8.46	7.89	2016.09	7.96	7.56	27.39
3000	-0.8		4.33	1023.79	3.57	3.02	7.67
	-0.5	5.90	110.52	22.89	11.98	8.89	1186.42
	0	1028.13	803.45	589.09	800.59	1025.26	1537.72
	0.5	24.54	1617.56	74.37	17.23	11.01	1653.84
	0.8	5.04	5.11	1651.30	3.70	3.09	16.79

Outline

- Correlated random vector generation
- MIP model
- Column generation
- Concluding remarks
- Future work
- References

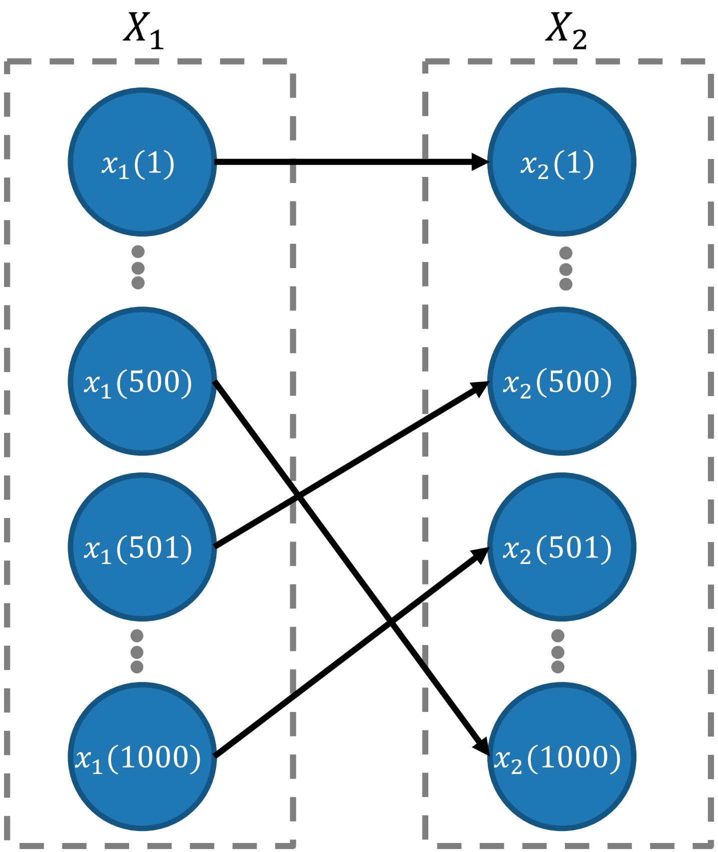
Concluding remarks

- We developed a distribution-free method to generate correlated variables.
- Better accuracy than the Iman & Conover method used in widely popular software like R, Crystal Ball, and @Risk.
- Challenge of the MIP-based method lies at the root node (linear relaxation).
- Column generation procedure accelerates the solution of the linear relaxation.
- The number of generated columns is always smaller than the sample size.
- We used warm initial assignments from Iman & Conover.
- Loss of quality of the solution (after enforcing integrality) is minimal.
- Significant speedup vs. pure MIP.
- The method behaves similarly for different mixtures of distributions (and different correlation targets).

Outline

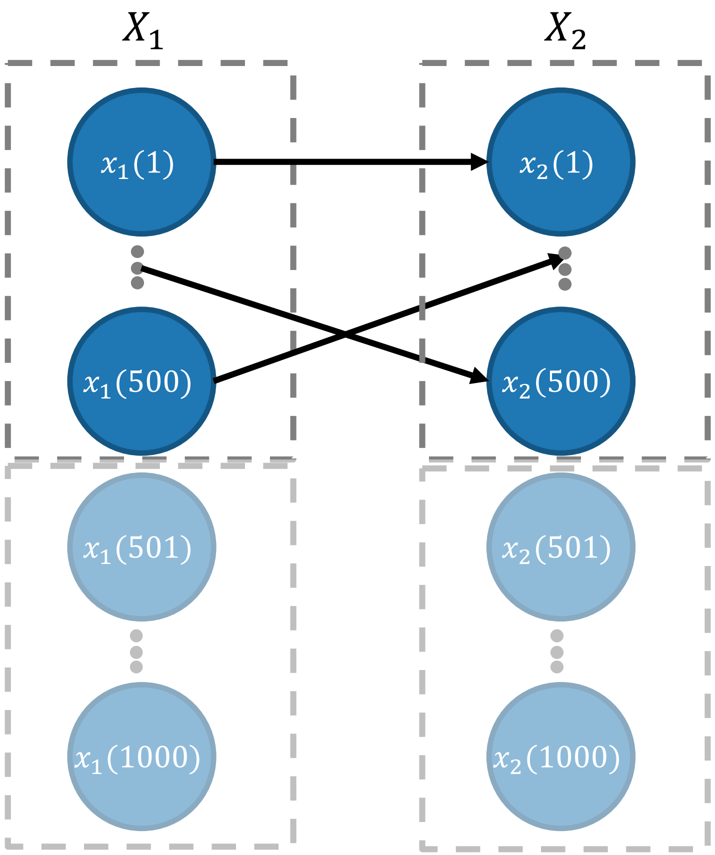
- Correlated random vector generation
- MIP model
- Column generation
- Concluding remarks
- Future work
- References

Future work
Simulation: Inducing correlation in batches

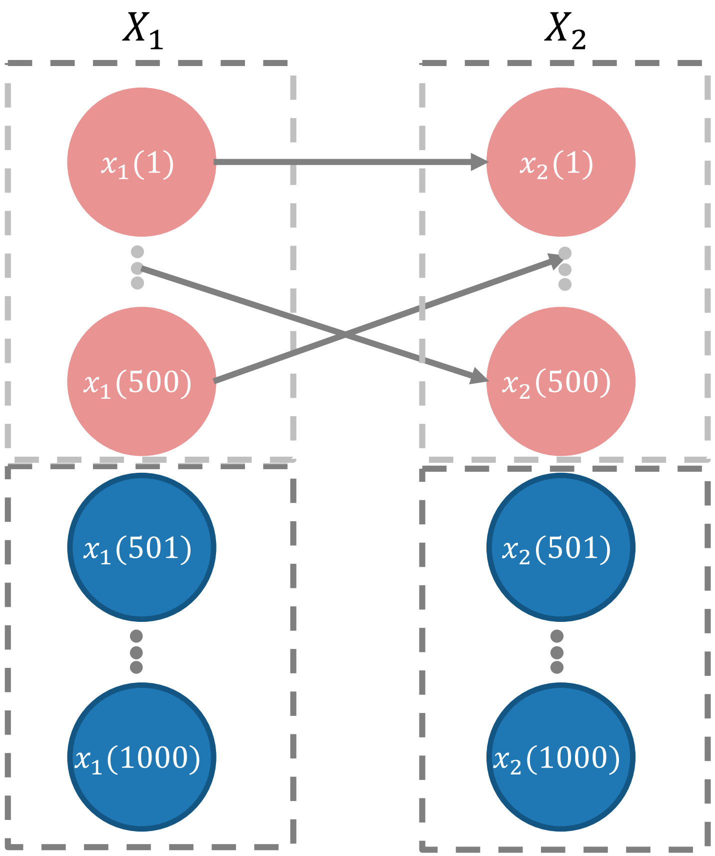


Future work

Simulation: Inducing correlation in batches



Future work
Simulation: Inducing correlation in batches



$$\min \delta^+ + \delta^-$$

s. t.

$$\frac{\sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij} z_{ij}}{S_{X_1 X_2}} + \delta^- - \delta^+ = \bar{\rho}^P_{X_1 X_2} + \alpha$$

Given by achieved correlation of fixed observations

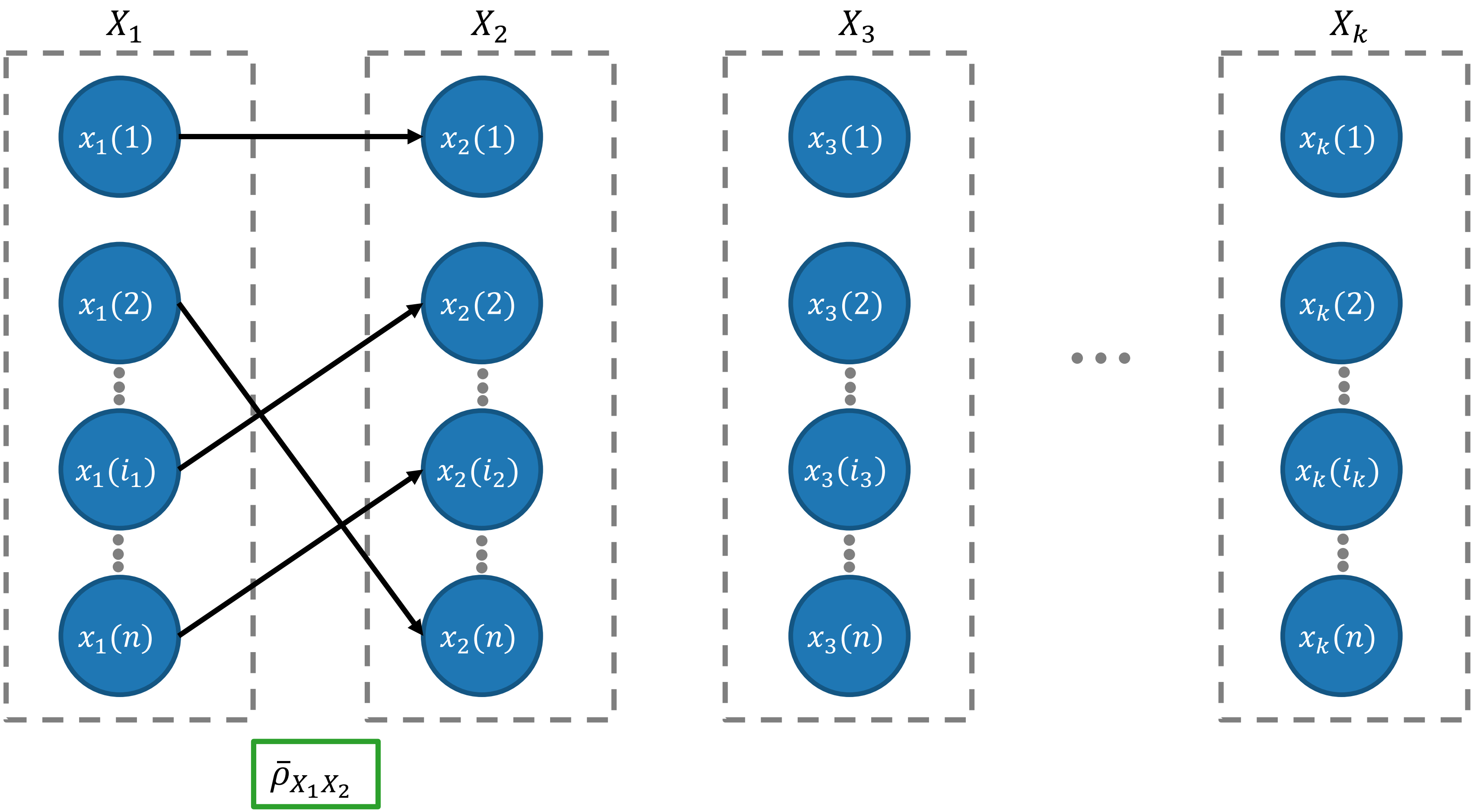
$$\sum_{j=1}^{n} z_{ij} = 1, \qquad i = 1, \dots, n$$

$$\sum_{i=1}^{n} z_{ij} = 1, \qquad j = 1, \dots, n$$

$$z_{ij} \geq 0, \qquad i, j = 1, \dots, n$$

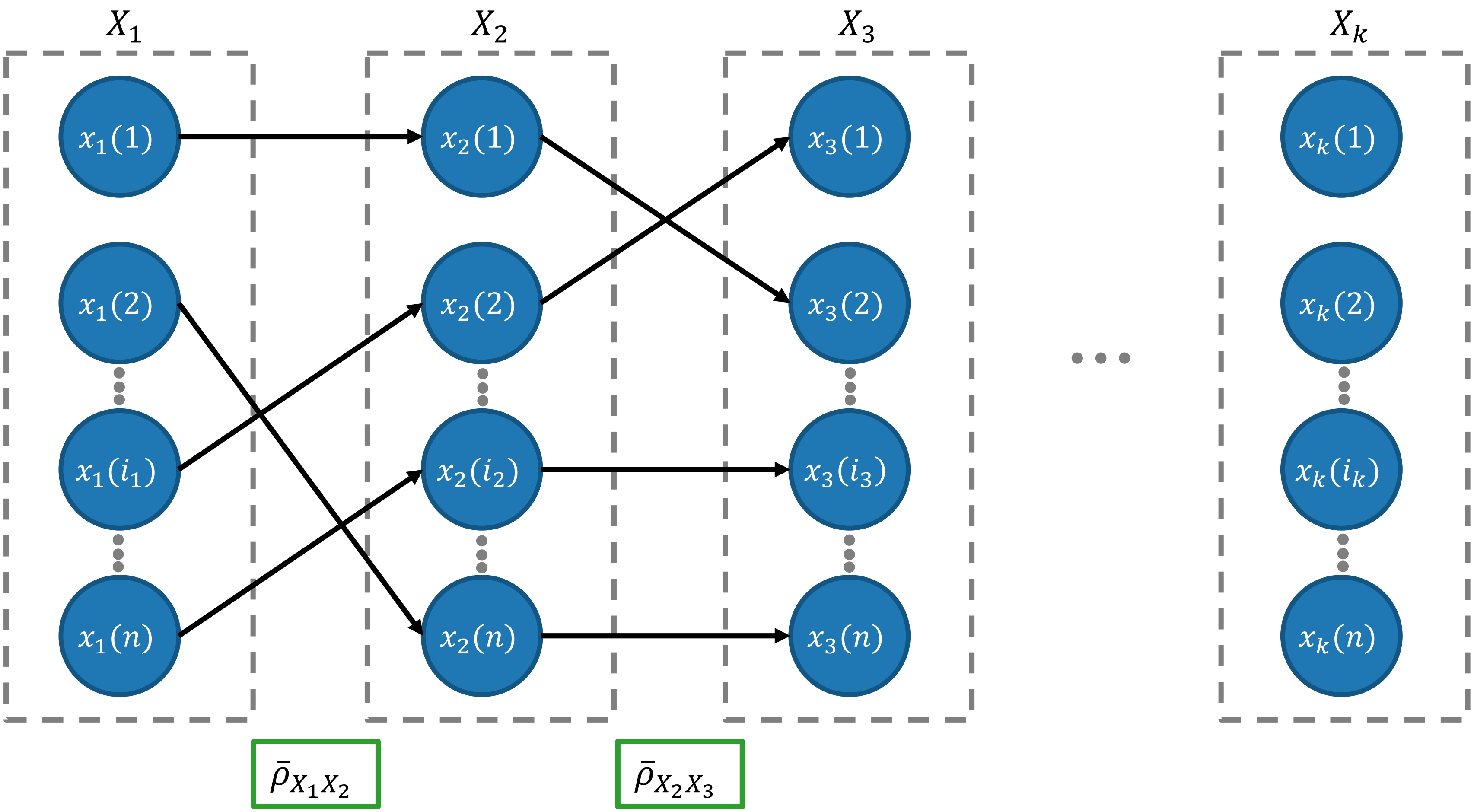
$$\delta^+, \delta^- \geq 0$$

Future work
Extension to multiple variables

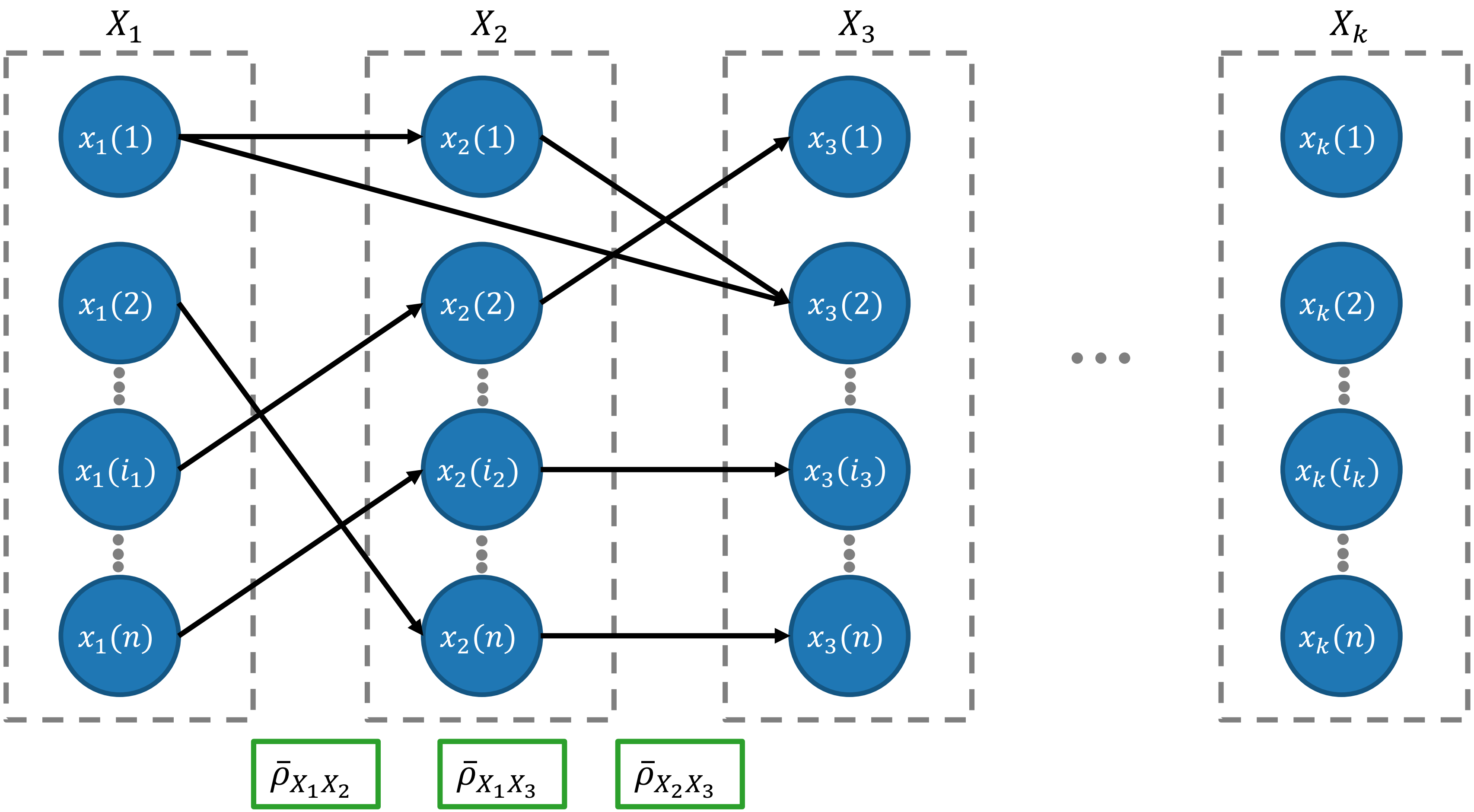


Universidad de los Andes

Future work
Extension to multiple variables



Future work
Extension to multiple variables



Universidad de los Andes

References

- Conway, D. A. (1979) Technical Report No. 145, Statistics Department, Stanford University.
- Iman, R. L. and Conover, W. J. (1982) A distribution-free approach to inducing rank correlation among input variables. *Communications in Statistics - Simulation and Computation*, 11(3):311-334.
- Law, A. M. and Kelton, W. D. (2000) Simulation Modeling and Analysis, 3rd ed., McGraw-Hill, New York, NY.
- Medaglia, A. L. and Sefair, J. A. (2009) Generating correlated random vectors using mixed-integer programming, in *Proceedings of the 2009 IIE Annual Conference*, Institute of Industrial Engineers (IIE), Miami, FL, pp. 1759-1764.
- Mitchell, C. R., Paulson, A. S. and Beswick, C. A. (1977) Effect of correlated exponential service times on single server tandem queues. Naval Research Logistics, 24(1):95-112.

Universidad de los Andes

Q&A

Thanks!

{amedagli, oo.guaje10, copa}@uniandes.edu.co
jorge.sefair@asu.edu

http://copa.uniandes.edu.co