Random function priors for exchangeable databases

James Robert Lloyd

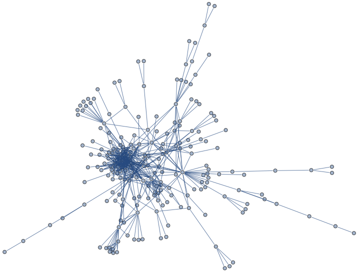
Machine Learning Group,
Department of Engineering,
University of Cambridge

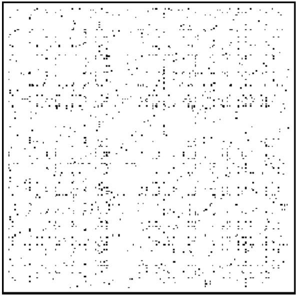
June 2013

Collaborators

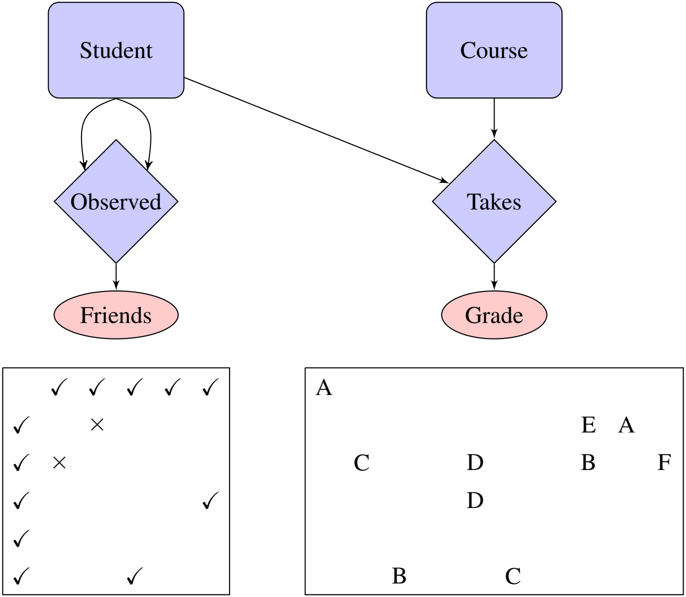
Daniel M. Roy (Cambridge)
Peter Orbanz (Columbia)
Zoubin Ghahramani (Cambridge)

Anything measured at more than one ‘object’

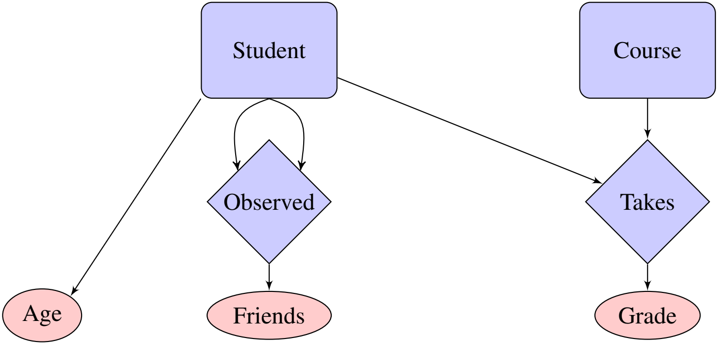




MULTIPLE RELATIONS



MULTIPLE RELATIONS AND COVARIATES / SEQUENCES



15
15
15
14
14
16

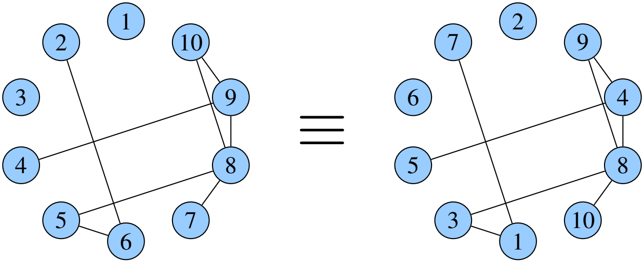
	✓	✓	✓	✓	✓
✓		×			
✓	×				
✓					✓
✓					
✓			✓		

A									
							E	A	
	C			D			B		F
				D					
		B			C				

Appropriate models for databases?

- Want to perform statistical tasks with this data
  - Predict unobserved data
  - Identify common structures e.g., group structure
- But what is an appropriate parameter space?
  - What is the target of statistical inference?
  - Where can we share statistical strength?
- What minimal assumptions can we make to answer these questions?

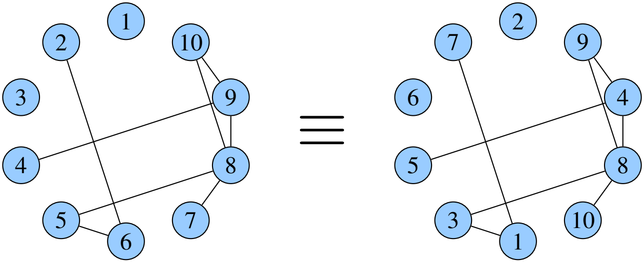
EXCHANGEABILITY

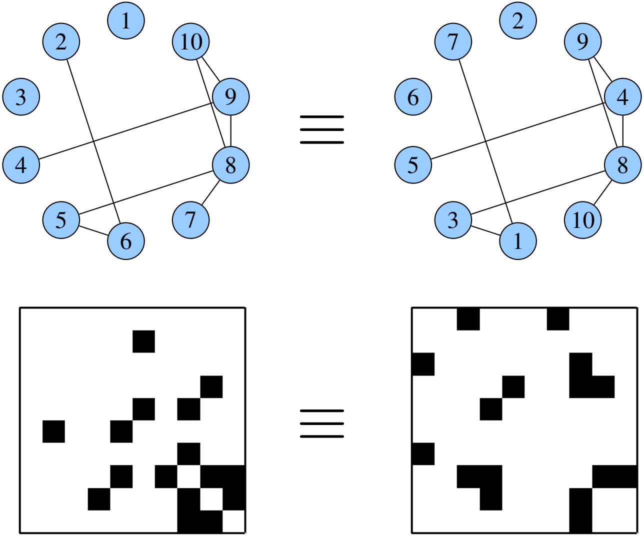


Overview

- Relational data typically encoded in arrays.
- Representation theorems for exchangeable arrays have previously been used to inspire Bayesian (non-parametric) modelling of relational data.
  - e.g., Eigenmodel [Hof08], Mondrian process [RT09], Random function model [LOGR12]
- We derive corollaries of these representation theorems applicable to collections of arrays i.e., databases…
- …and discuss implications for modelling such data

Exchangeability for relational data




1
2
10
3
9
4
8
5
7
6
≡
2
7
9
6
4
5
8
3
10
1
≡

Exchangeable array representation

Definition

A d-array $X = (X_{i_1 \dots i_d})_{i_j \in \mathbb{N}}$ is called a (jointly/weakly) *exchangeable array* if

$$(X_{i_1 \dots i_d}) \overset{\mathrm{d}}{=} (X_{p(i_1) \dots p(i_d)}) \qquad \text{for every } p \in \mathbb{S}_\infty \, .$$

Theorem (Aldous [Ald81], Hoover [Hoo82])

A random 2-array (X_{ij}) is exchangeable if and only if there exists a random measurable function $F : [0,1]^3 \to \mathcal{X}$ such that

$$(X_{ij}) \overset{d}{=} (F(U_i, U_j, U_{ij})).$$

where $(U_i)_{i \in \mathbb{N}}$ and $(U_{ij})_{i \leq j \in \mathbb{N}}$ are i.i.d. $\mathrm{Uniform}[0,1]$ *random variables and $U_{ji} = U_{ij}$ for $j < i \in \mathbb{N}$.*

Definition

A d-array $X = (X_{i_1 \dots i_d})_{i_j \in \mathbb{N}}$ is called a (jointly/weakly) *exchangeable array* if

$$(X_{i_1 \dots i_d}) \overset{\mathrm{d}}{=} (X_{p(i_1) \dots p(i_d)}) \qquad \text{for every } p \in \mathbb{S}_\infty \ .$$

Theorem (Aldous [Ald81], Hoover [Hoo82])

A random 2-array (X_{ij}) is exchangeable if and only if there exists a measurable function $F : [0,1]^4 \to \mathcal{X}$ such that

$$(X_{ij}) \overset{d}{=} (F(\alpha, U_i, U_j, U_{ij})).$$

where α, $(U_i)_{i \in \mathbb{N}}$ and $(U_{ij})_{i \leq j \in \mathbb{N}}$ are i.i.d. $\text{Uniform}[0,1]$ *random variables and $U_{ji} = U_{ij}$ for $j < i \in \mathbb{N}$.*

Theorem (3-arrays)

A random 3-array (X_{ijk}) is exchangeable if and only if there exists a measurable function $F : [0,1]^8 \to \mathcal{X}$ such that

$$(X_{ijk}) \overset{d}{=} (F(\alpha, U_i, U_j, U_k, U_{ij}, U_{ik}, U_{jk}, U_{ijk})).$$

Theorem (4-arrays)

A random 4-array (X_{ijkl}) is exchangeable if and only if there exists a measurable function $F : [0,1]^{15} \to \mathcal{X}$ such that

$$(X_{ijkl}) \overset{d}{=} (F(\alpha, U_i, U_j, U_k, U_l, U_{ij}, U_{ik}, U_{il}, U_{jk}, U_{jl}, U_{kl}, U_{ijk}, U_{ijl}, U_{jkl}, U_{ijkl})).$$

Theorem (d-arrays)

New notation required - see [Kal99]

A simpler representation can be used

Call a d-array $(X_{i_1 \dots i_d})$ *simple* if it admits a representation

$$(X_{i_1 \dots i_d}) \overset{d}{=} (\Theta(U_{i_1}, \dots, U_{i_d}))$$

where $\Theta : [0,1]^d \to \mathcal{X}$ is a random measurable function and $(U_i)_{i \in \mathbb{N}}$ are i.i.d. Uniform$[0,1]$ random variables.

Theorem (Kallenberg [Kal99])

Let X be an exchangeable array in a Borel space $\mathcal{X}$. Then there exist some simple exchangeable arrays $X_1, X_2, \dots$ such that $\mathcal{L}(\chi_m X_n)$ and $\mathcal{L}(\chi_m X)$ are mutually absolutely continuous for all $m, n \in \mathbb{N}$ and the associated Radon–Nikodym derivatives converge uniformly to 1 as $n \to \infty$ for fixed m.

Notation

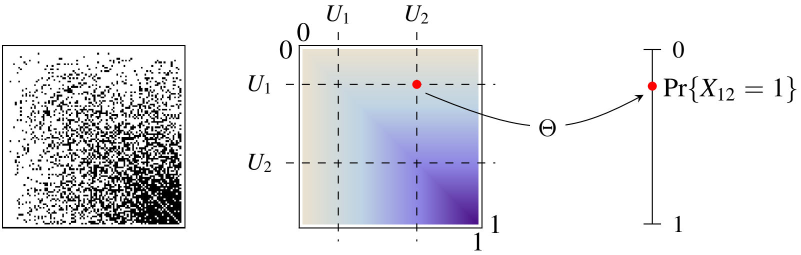
$\mathcal{L}(Y)$ is the law (distribution) of a random variable Y and $\chi_m X := (X_{i_1 \dots i_d};\ i_j \leq m)$.
We denote the type of convergence above by $\overset{d}{\rightsquigarrow}$.

Representation results inspire a generic modelling recipe

e.g., Binary networks

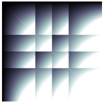
Θ	-	Adjacency matrix approximated by function on unit square
U_i	-	Each node associated with a latent variable in $[0, 1]$
$W_{ij} := \Theta(U_i, U_j)$	-	Evaluation of approximate adjacency matrix
$X_{ij} \sim \text{Bernoulli}(W_{ij})$	-	Bernoulli likelihood (can be shown to be general)

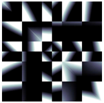
Θ can be pictured as a blurred adjacency matrix



Eigenmodel [Hof08]

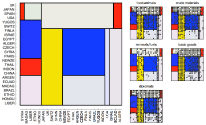
$$U_i \sim_{\text{iid}} \mathcal{N}(0, I)$$
$$\Theta \sim \text{Bilinear}$$





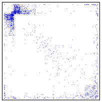
The Mondrian Process [RT09]

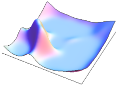
$$U_i \sim_{\text{iid}} \text{Uniform}[0, 1]$$
$$\Theta \sim \text{Piecewise constant}$$



The Random Function Model [LOGR12]

$$U_i \sim_{\text{iid}} \mathcal{N}(0, I)$$
$$\Theta \sim \mathcal{GP}(0, \kappa)$$





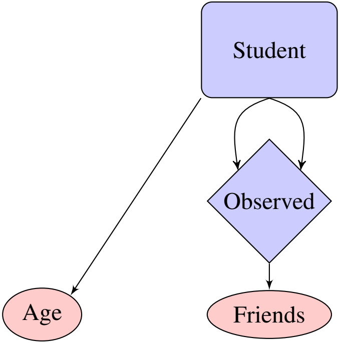
Many other models fit this pattern

	W_{ij}	κ	$U_i \sim$.
Random function model	$\phi(U_i)'\Lambda$	$\kappa_{U\times U}$	Gaussian
SMGB, InfTucker	$\phi(U_i)'\Lambda\phi(U_j)$	$\kappa_U \otimes \kappa_U$	Laplace
GPLVM	$\phi(U_i)'\Lambda$	$\kappa_U \otimes \delta$	Gaussian
Eigenmodel	$U_i'\Lambda U_j$	$L_U \otimes L_U$	Gaussian
Linear relational GP	$U_i'\Lambda U_j$	$L_U \otimes L_U$	Gaussian
PCA, PMF	$U_i'\Lambda$	$L_U \otimes \delta$	Gaussian
Latent distance	$-\lvert U_i - U_j\rvert$	*	Gaussian
Mondrian process based	Decision tree	*	Uniform
Latent class	$\Lambda_{U_i U_j}$	$\delta_{U\times U}$	Multinomial
IRM, IHRM	$\Lambda_{U_i U_j}$	$\delta_{U\times U}$	CRP
BMF, LFRM	$U_i'\Lambda U_j$	$L_U \otimes L_U$	IBP
ILA	$\sum_d \mathbb{I}_{U_{id}} \mathbb{I}_{U_{jd}} \Lambda^{(d)}_{U_{id}U_{jd}}$	*	CRP + IBP

Notes

κ is the kernel in the often equivalent Gaussian process representation; ϕ is the corresponding feature map. L is a linear kernel, δ is the Kronecker delta function, $\otimes$ is a tensor / Kronecker product. Λ is a matrix. $\mathbb{I}$ is an indicator function.

WHAT ABOUT COVARIATES?



15
15
15
14
14
16

	✓	✓	✓	✓	✓
✓		×			
✓	×				
✓					✓
✓					
✓			✓		

Suppose that in addition to a social network (X_{ij}) we have side information in the form of covariates for the users (C_i).

Corollary

Let $(X_{ij})_{i,j \in \mathbb{N}}$ *and* $(C_i)_{i \in \mathbb{N}}$ *be random variables in* $\mathcal{X}$ *and* $\mathcal{X}'$ *respectively. Then the following are equivalent:*

i. $(X_{ij}), (C_i) \stackrel{d}{=} (X_{p(i)p(j)}), (C_{p(i)})$ *for every* $p \in \mathbb{S}_\infty$.

ii. *There are random measurable functions* $F : [0,1]^3 \to \mathcal{X}$ *and* $G : [0,1] \to \mathcal{X}'$ *such that*

$$(X_{ij}), (C_i) \stackrel{d}{=} (F(U_i, U_j, U_{ij})), (G(U_i)),$$

where $(U_i)_{i \in \mathbb{N}}$ *and* $(U_{ij})_{i \leq j \in \mathbb{N}}$ *are i.i.d.* $\mathrm{Uniform}[0,1]$ *random variables and* $U_{ji} = U_{ij}$ *for* $j < i \in \mathbb{N}$.

Proof sketch

Let $(R_{ij}) := ((X_{ij}, C_i))$. This array is jointly / weakly exchangeable.

By Aldous–Hoover, there exists a measurable function F' into $(\mathcal{X}, \mathcal{X}')$ such that

$$(R_{ij}) := ((X_{ij}, C_i)) \stackrel{\mathrm{d}}{=} (F'(\alpha, U_i, U_j, U_{ij})),$$

and so we may write

$$(X_{ij}, C_i) \stackrel{\mathrm{d}}{=} (F'_1(\alpha, U_i, U_j, U_{ij}), F'_2(\alpha, U_i, U_j, U_{ij}))$$

for a pair of measurable functions F'_1 and F'_2 into $\mathcal{X}$ and $\mathcal{X}'$, respectively.

We remove the redundant representations for C_i by defining

$$G(\alpha, U_i) = \int F'_2(\alpha, U_i, x, y)\,\mathrm{d}x\,\mathrm{d}y.$$

The proof concludes by demonstrating the required properties of G.

Corollary (Simple array approximation)

Let $(X_{ij})_{i,j\in\mathbb{N}}$ *and* $(C_i)_{i\in\mathbb{N}}$ *be random variables in* $\mathcal{X}$ *and* $\mathcal{X}'$ *respectively. Then i) implies ii):*

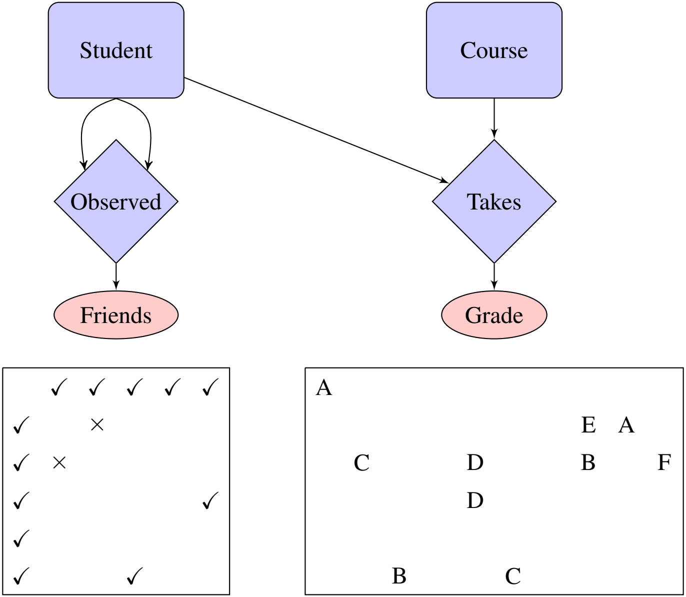
i. $(X_{ij}), (C_i) \overset{d}{=} (X_{p(i)p(j)}), (C_{p(i)})$ *for every* $p \in \mathbb{S}_\infty$.

ii. *There is a sequence of random measurable functions* $F^n : [0,1]^2 \to \mathcal{X}$ *and* $G^n : [0,1] \to \mathcal{X}'$ *such that*

$$(F^n(U_i, U_j)), (G^n(U_i)) \overset{d}{\rightsquigarrow} (X_{ij}), (C_i)$$

where $(U_i)_{i\in\mathbb{N}}$ *are i.i.d.* Uniform$[0,1]$ *random variables* .

WHAT ABOUT MULTIPLE RELATIONS?



Consider rating data (X_{ij}) with users i and items j, and a social network (S_{ik}) over users i, k.

Corollary

The following are equivalent

i. $(X_{ij}), (S_{ik}) \overset{d}{=} (X_{p(i)p'(j)}), (S_{p(i)p(k)})$ *for every* $p, p' \in \mathbb{S}_\infty$.

ii. *There exist random functions* F, G *such that*

$$(X_{ij}), (S_{ik}) \overset{d}{=} (F(U_i, V_j, W_{ij})), (G(U_i, U_k, U_{ik}))$$

where $(U_i)_{i\in\mathbb{N}}, (V_j)_{j\in\mathbb{N}}, (W_{ij})_{i,j\in\mathbb{N}}$ *and* $(U_{ik})_{i\leq k\in\mathbb{N}}$ *are i.i.d.* Uniform$[0, 1]$ *random variables, and* $U_{ki} = U_{ik}$ *for* $k < i \in \mathbb{N}$.

EXTENSIONS: TWO ARRAYS WITH FEATURE DATA

Consider rating data (X_{ij}) with users i and items j, with side information in the form of covariates for both users, C_i, and movies, D_j, and a social network (S_{ik}) over users i, k.

Corollary

The following are equivalent

i. $(X_{ij}), (C_i), (D_j), (S_{ik}) \stackrel{d}{=} (X_{p(i)p'(j)}), (C_{p(i)}), (D_{p'(j)}), (S_{p(i)p(k)})$ *for every* $p, p' \in \mathbb{S}_\infty$.

ii. *There exist random functions* F, G, H, I *such that*

$$(X_{ij}), (C_i), (D_j), (S_{ik}) \stackrel{d}{=} (F(U_i, V_j, W_{ij})), (G(U_i)), (H(V_j)), (I(U_i, U_k, U_{ik}))$$

where $(U_i)_{i\in\mathbb{N}}, (V_j)_{j\in\mathbb{N}}, (W_{ij})_{i,j\in\mathbb{N}}$ *and* $(U_{ik})_{i\leq k\in\mathbb{N}}$ *are i.i.d.* Uniform$[0, 1]$ *random variables, and* $U_{ki} = U_{ik}$ *for* $k < i \in \mathbb{N}$.

Corollary (Simple array approximation)

i) implies ii)

i. $(X_{ij}), (S_{ik}) \overset{d}{=} (X_{p(i)p'(j)}), (S_{p(i)p(k)})$ *for every* $p, p' \in \mathbb{S}_\infty$.

ii. *There exists a sequence of random functions* F^n, G^n *such that*

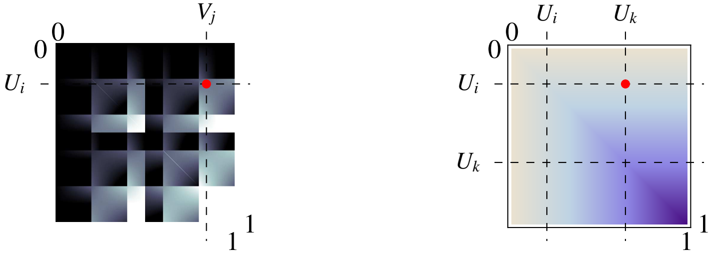
$$(F^n(U_i, V_j)), (G^n(U_i, U_k)) \overset{d}{\rightsquigarrow} (X_{ij}), (S_{ik})$$

where $(U_i)_{i \in \mathbb{N}}, (V_j)_{j \in \mathbb{N}}$ *are* $i.i.d.$ $\mathrm{Uniform}[0,1]$ *random variables.*

Modelling recipe extends to multiple arrays

F, G	-	Functions on unit square
U_i, V_j	-	Each object associated with a latent variable
$V_{ij}, W_{ik} := F(U_i, V_j), G(U_i, U_k)$	-	Evaluation of array approximations
$X_{ij} \sim \text{Poisson}(V_{ij})$	-	e.g., Poisson likelihood
$S_{ik} \sim \text{Bernoulli}(W_{ik})$	-	e.g., Bernoulli likelihood

Two arrays modelled by functions F and G



Suppose we have a database consisting of R relations across O objects. Let m^r_o be the number of times relation r uses object o as an input i.e., relation r is a function with domain $\mathbb{N}^{m^r_1} \times \ldots \times \mathbb{N}^{m^r_O}$.

Encode relation r in an array

$$(X^r) = (X^r_{i_1,\ldots,i_{m^r_1},i_{m^r_1+1},\ldots,i_{m^r_1+m^r_2},\ldots,i_{\sum_{o=1}^{O-1} m^r_o+m^r_O}})_{i_j\in\mathbb{N}}.$$

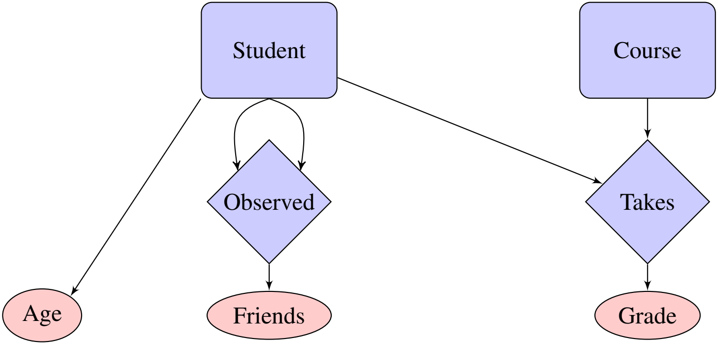
If the ordering of all objects is arbitrary, then X^r is π-exchangeable where π is the partition of consecutive integers with lengths $m^r_1, m^r_2, \ldots, m^r_O$.

Corollary

There exists a sequence of random measurable functions (F^r_n) such that

$$(F^r_n(U^1_{i_1},\ldots,U^1_{i_{m^r_1}},U^2_{i_{m^r_1+1}},\ldots,U^2_{i_{m^r_1+m^r_2}},\ldots,U^O_{i_{\sum_{o=1}^{O-1} m^r_o+m^r_O}})) \overset{d}{\rightsquigarrow} (X^r) \quad \forall\, r$$

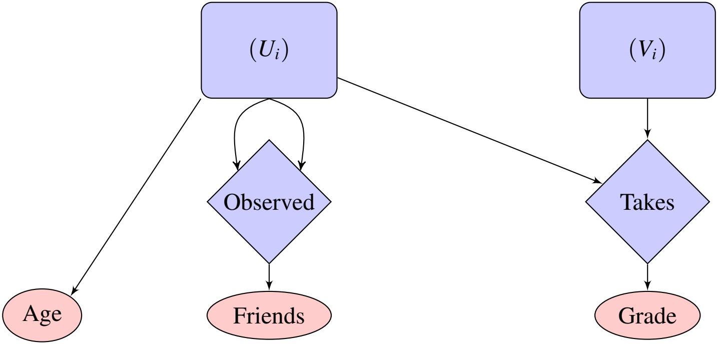
where $(U^o_{i_j})$ are i.i.d. uniform random variables.



15
15
15
14
14
16

	✓	✓	✓	✓	✓
✓		×			
✓	×				
✓					✓
✓					
✓			✓		

A									
							E	A	
	C			D			B		F
				D					
		B			C				

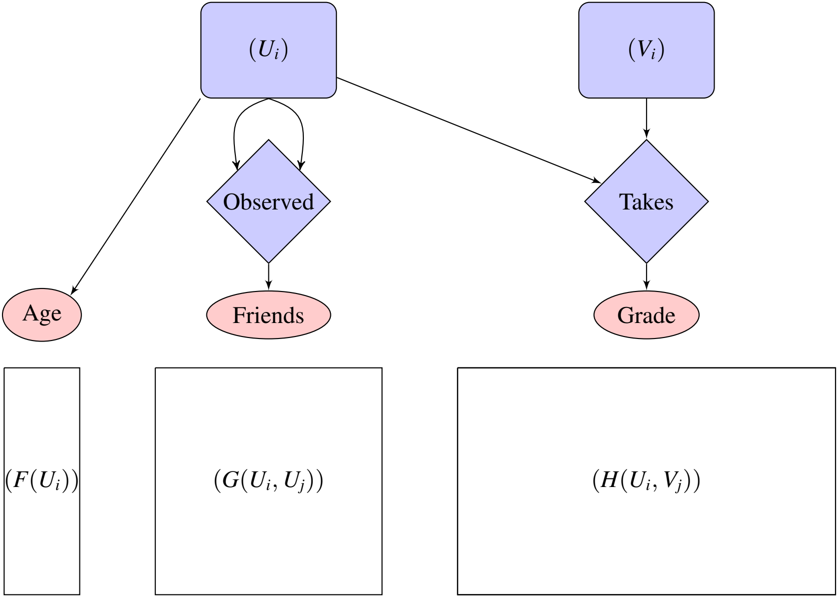


15
15
15
14
14
16

	✓	✓	✓	✓	✓
✓		×			
✓	×				
✓					✓
✓					
✓			✓		

A									
							E	A	
	C			D			B		F
				D					
		B			C				

AND LATENT FUNCTIONS FOR EACH ARRAY



Inifinite relational model [KTG+06]

- Objects share partition structure across different relations
- Independent class interaction probabilities

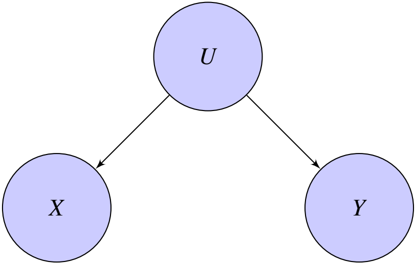
Other examples

- Coupled Tensor Factorisation [YCS11]
- Compiling Relational Database Schemata into Probabilistic Graphical Models [SG12]

Note about function independence

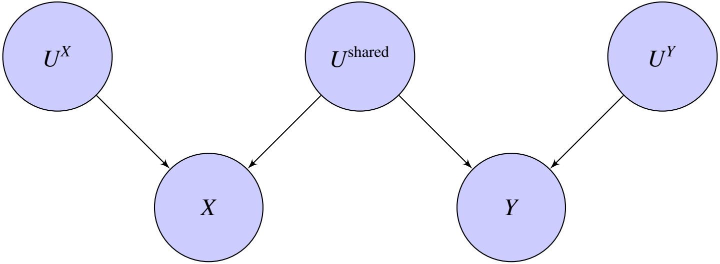
- All models assume a priori independence of the representing functions
- In general, the functions may be dependent
- May result in statistically inefficient modelling

Modelling practicalities



- Sharing latent variables between arrays may result in poor modelling compromise

Modelling practicalities



- May be alleviated by allowing model to flexibly use different dimensions / parts of latent variables for different arrays (c.f. automatic relevance determination)
- Non-parametric approach may be particularly helpful, allowing dimensionality of latent variables to grow when modelling several arrays

SUMMARY

- Relational data can be represented by (collections of) arrays i.e., databases
- When these arrays can be assumed to be exchangeable...
  - Applicable when ordering of data is unimportant or arbitrary
- ...their distributions can be characterised by corollaries of representation theorems for single arrays
- Representations reveal an appropriate parameter space, elucidating the targets of statistical inference

REFERENCES I

[Ald81] David J. Aldous. Representations for partially exchangeable arrays of random variables. *J. Multivariate Anal.*, 11(4):581–598, 1981.

[Hof08] Peter D. Hoff. Modeling homophily and stochastic equivalence in symmetric relational data. In *Advances in Neural Information Processing Systems (NIPS)*, volume 20, pages 657–664, 2008.

[Hoo82] D N Hoover. Row-column exchangeability and a generalized model for probability. In *Exchangeability in Probability and Statistics*, pages 281–291, 1982.

[Kal99] Olav Kallenberg. Multivariate sampling and the estimation problem for exchangeable arrays. *Journal of Theoretical Probability*, 12(3):859–883, 1999.

[KTG+06] Charles Kemp, Joshua B Tenenbaum, Thomas L Griffiths, Takeshi Yamada, and Naonori Ueda. Learning systems of concepts with an infinite relational model. In *Proceedings of the national conference on artificial intelligence*, volume 21, page 381. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; 1999, 2006.

[LOGR12] James Robert Lloyd, Peter Orbanz, Zoubin Ghahramani, and Daniel M. Roy. Random function priors for exchangeable arrays with applications to graphs and relational data. In *Advances in Neural Information Processing Systems*, 2012.

[PKG12] Konstantina Palla, David A. Knowles, and Zoubin Ghahramani. An infinite latent attribute model for network data. In *Proceedings of the 29th International Conference on Machine Learning*, ICML 2012. Edinburgh, Scotland, GB, July 2012.

[RT09] D. M. Roy and Y. W. Teh. The Mondrian process. In *Advances in Neural Information Processing Systems*, volume 21, 2009.

[SG12] Sameer Singh and Thore Graepel. Compiling relational database schemata into probabilistic graphical models. *arXiv preprint arXiv:1212.0967*, 2012.

[YCS11] Y Kenan Yılmaz, A Taylan Cemgil, and Umut Simsekli. Generalised coupled tensor factorisation. In *Neural Information Processing Systems*, 2011.