

copula and vine

Fedde Tolman

fedde.tolman@kiwa.com

DACE-PRA 27-09-2018

doel sig-PRA

- werkbare methode voor PRA
 - data -> acceptabele hoeveelheid werk vs. nauwkeurigheid en betrouwbaarheid
eerdere lezingen
 - model -> veel en uitgebreide software, dus er worden veel fouten gemaakt -> begrip van beginselen nodig
onderwerp deze presentatie
- relevantie van PRA
 - kosten van risicoanalyse en -berekening vs. opbrengst
voornemen

stochast

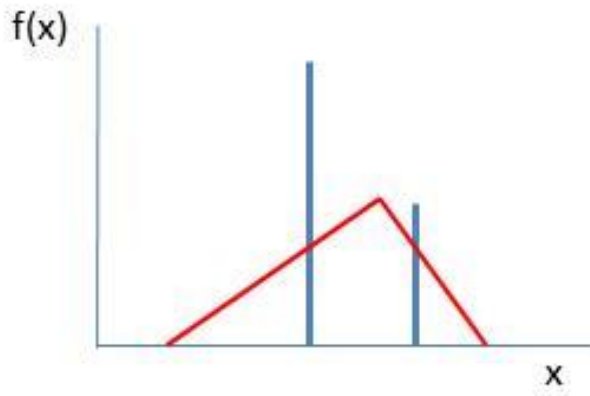
- een stochastische variabele is een grootheid waarvan de waarde afhangt van toeval

*(G. stokhos = guess, aim, fixed target, erected pillar for archers to shoot at; PIE *stogh-, stegh- = to stick, prick, sting)*

- beschrijving door
 - kansfunctie
 - dichtheid $f(x)$
 - verdeling $F(x)$
 - parameters (momenten) van een kansfunctie
 - gemiddelde μ / standaardafwijking σ / scheefheid / spitsheid, ...

meten of schatten van een stochast

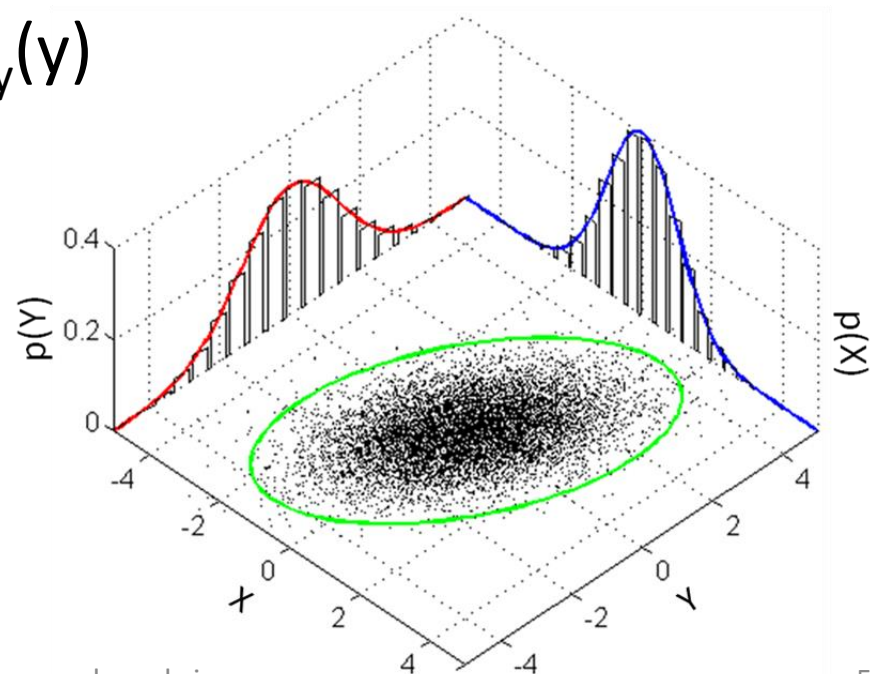
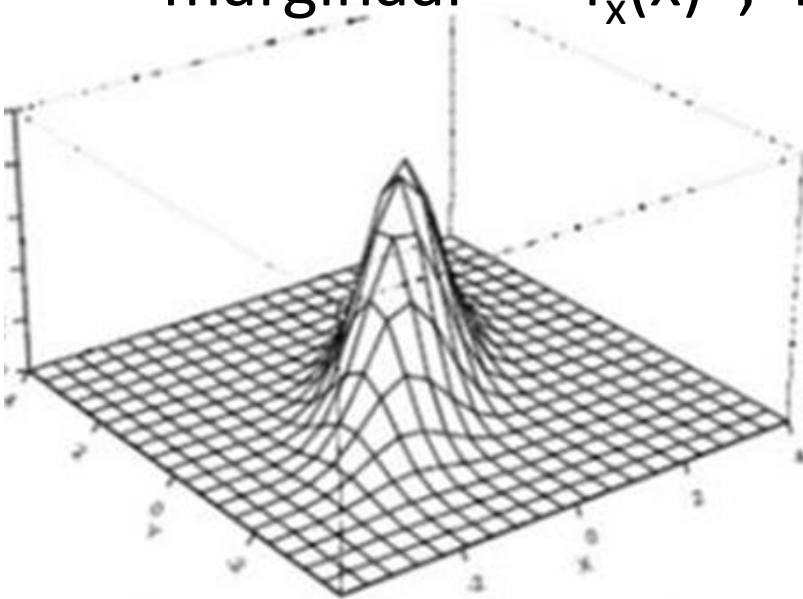
- binair (twee-waardig)
 - minimum – maximum
 - f discreet $\rightarrow$ p schatten
- triangulair (drie-waardig)
 - minimum – modus – maximum
 - aanname: f driehoeksverdeling $\rightarrow$ f hoeft niet geschat te worden



beschrijving	binair	triangle
variabelen	$x_1, x_2,$	x_1, x_2, x_3
kans	p	
parameters	μ, σ	μ, σ

verdelingen van twee stochasten

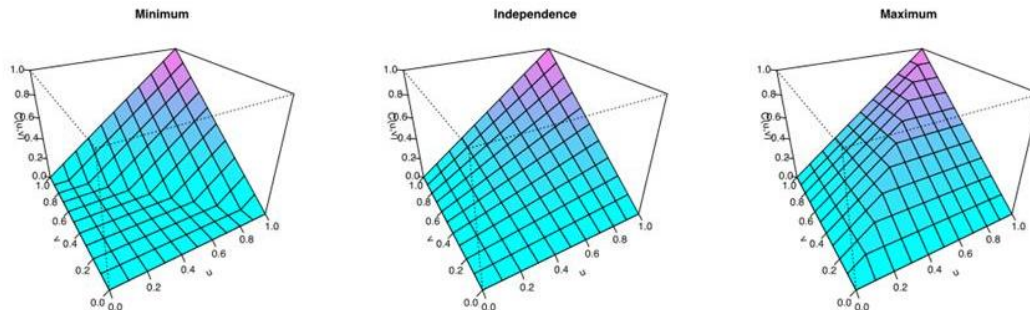
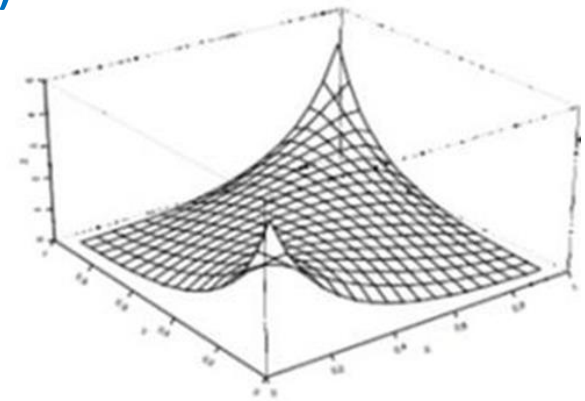
- gezamenlijk (simultaan) $f(x, y)$
- partieel
 - conditioneel $f_x(x|y)$, $f_y(y|x)$
 - marginaal $f_x(x)$, $f_y(y)$



onderling verband tussen stochasten

copula C

- twee stochasten $F(x,y) = C(F_x(x), F_y(y))$
 $f(x,y) = c(f_x(x), f_y(y))$
- meerdere stochasten $F(x_i, \dots) = C(F_i(x_i))$
 $f(x_i, \dots) = c(f_i(x_i))$
- C is vaak een moeilijke functie
- bepaalbare onder- en bovengrenzen (Fréchet–Hoeffding) -> iteratieve convergentie



co-relatie ρ

- vereenvoudiging van copula: correlatie (soort parameter van de copula-functie)
- in principe enkel voor $N(\mu_x, \sigma_x, \mu_y, \sigma_y, \rho_{xy})$; andere verdelingen hebben andere parameters voor een x-y verband
- de correlatiecoëfficiënt ρ is de sterkte van dat verband

PRA stelling 1:

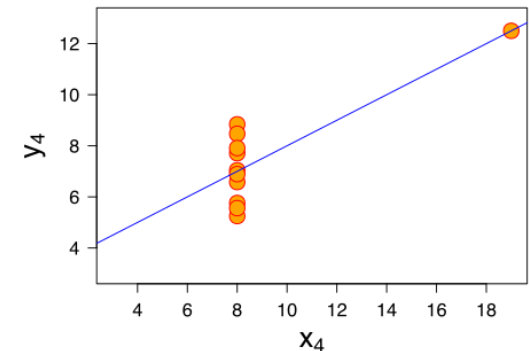
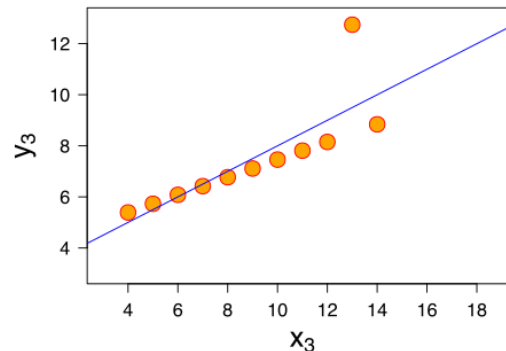
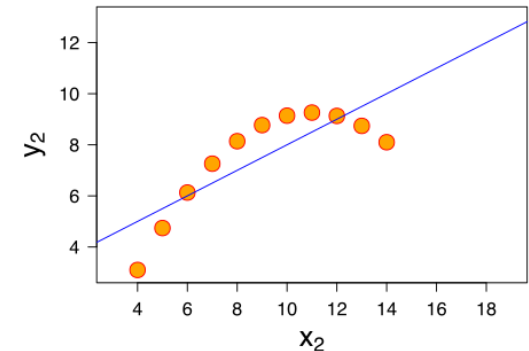
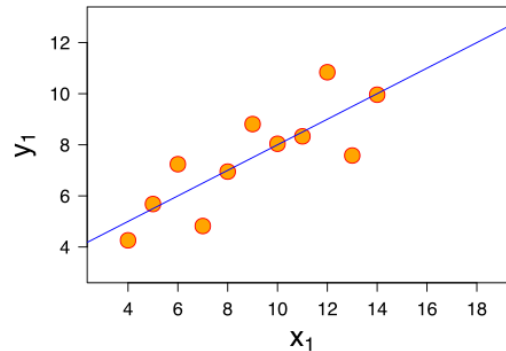
Point Estimation Method

- (Emilio) Rosenblueth (1975)
- per 2 variabelen – 3^2 waarden (i.p.v. 2×3 waarden)
- volledige correlatiematrix vergt echter $\frac{1}{2}n(n-1)$ paren

f (x, y)	y1	y2	y3	f _y (y)	f _. (.)	μ	σ	ρ
x1					x			
x2					y			
x3								
f _x (x)								

ρ ogenschijnlijk eenvoudiger maar lastiger te begrijpen dan c

- voorbeeld
 $y = 3.0 + 0.5x$
 $\rho = 0.816$
(Anscombe, 1973)



- 13 definities (Rodgers, Nicewander, The American Statistician, 1988)

correlation matrix P

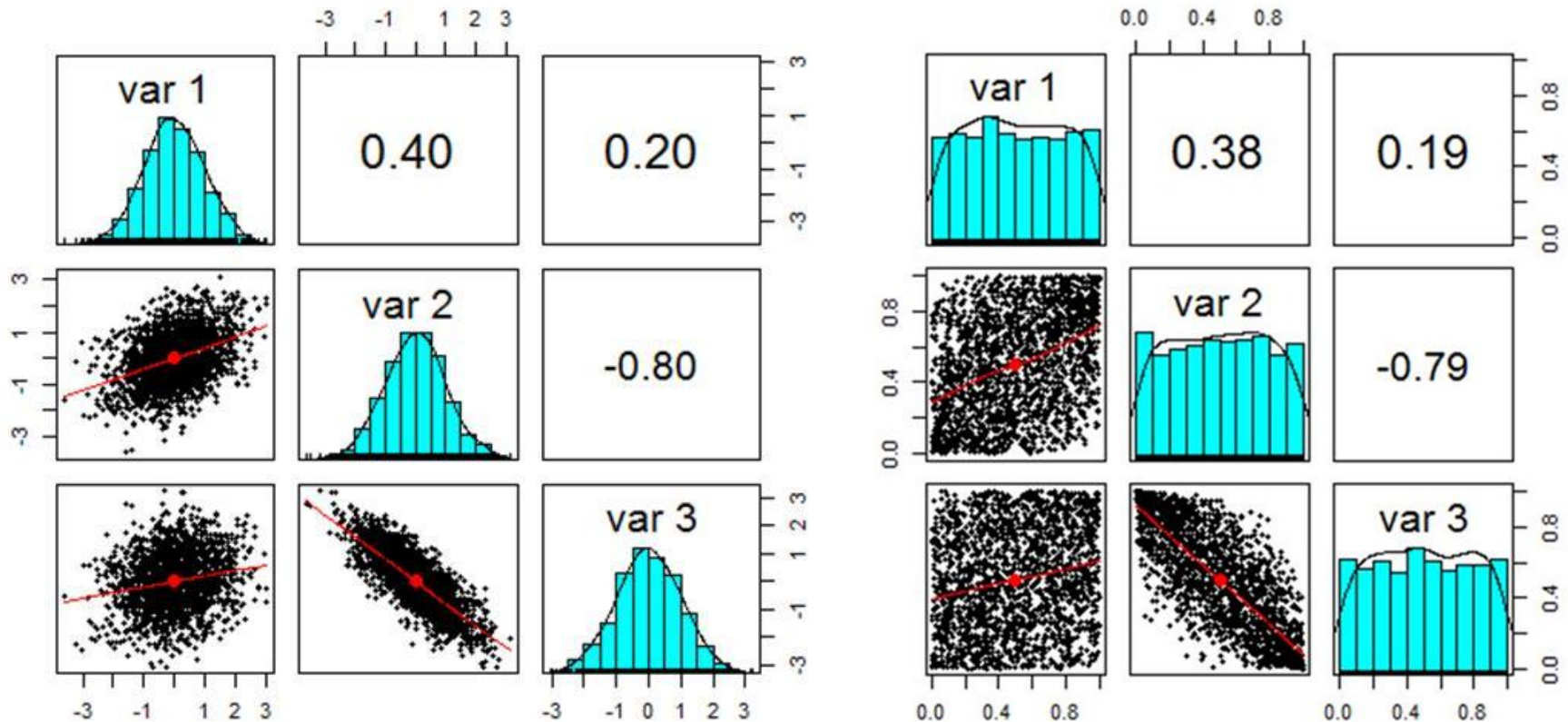
- n stochastics
- n^2 matrix P

requirements:

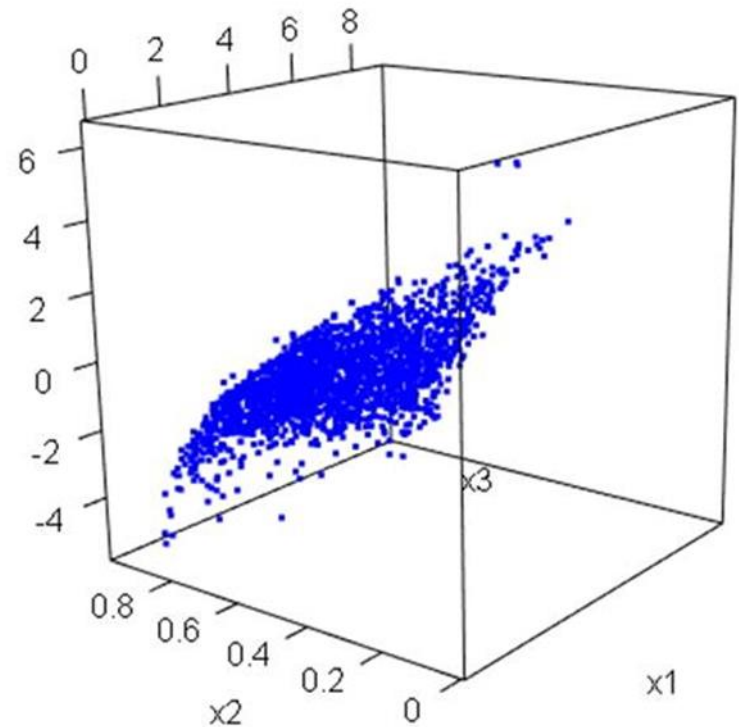
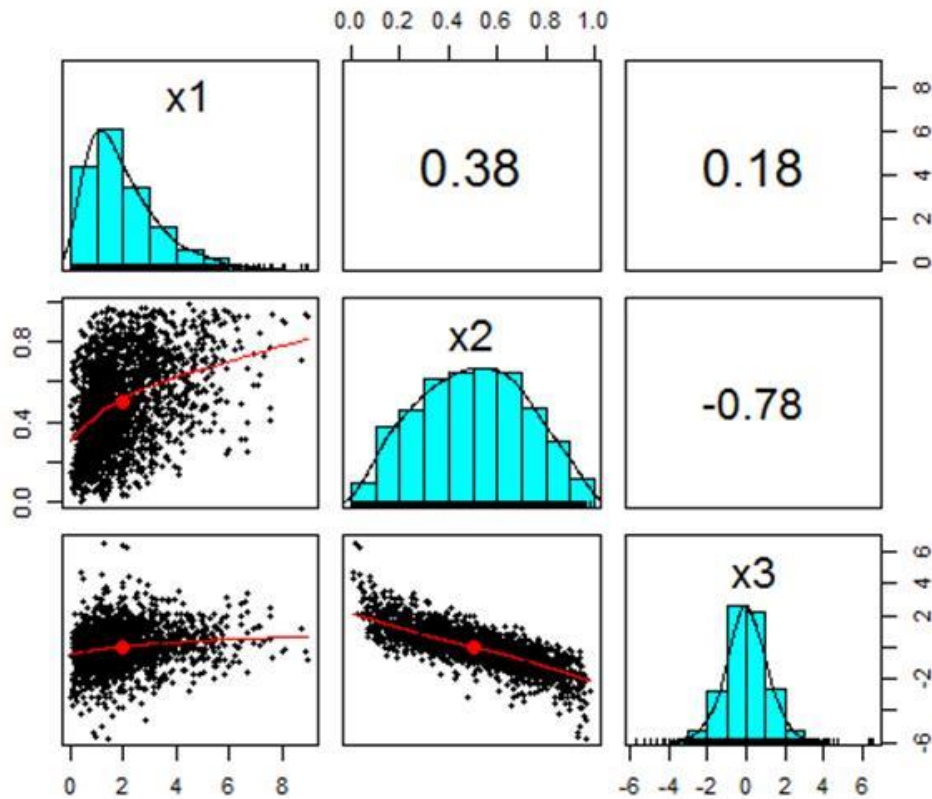
- $-1 \leq \rho_{ij} \leq 1$
- symmetrisch $\rightarrow \frac{1}{2} n(n-1)$ cc
- P is positief definite (*4 equivalent definitions*)
 - $x^T R x > 0$ for any $x \neq 0$.
 - R has positive eigenvalues
 - every principal determinant is positive (Sylvester)
 - positive pivots in Gaussian elimination

$$\begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{21} & 1 & \cdots & \rho_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{n1} & \cdots & \cdots & 1 \end{bmatrix}$$

normale en uniforme verdelingen



verschillende verdelingen



triangulatie

partiele correlatie coefficient

3 variabelen, 3 relaties -> logische consistentie

- $A > B, B > C \rightarrow A > C$ (niet vrij te kiezen)
- maar ook: $A > B, A > C \rightarrow A < C$ ligt niet vast

$$\begin{bmatrix} 1 & a & c \\ a & 1 & b \\ c & b & 1 \end{bmatrix}$$

$$ab - \sqrt{(1 - a^2)(1 - b^2)} \leq c \leq ab + \sqrt{(1 - a^2)(1 - b^2)}$$

$$\rho_{xz|y} = \frac{c - ab}{\sqrt{(1 - a^2)(1 - b^2)}}$$

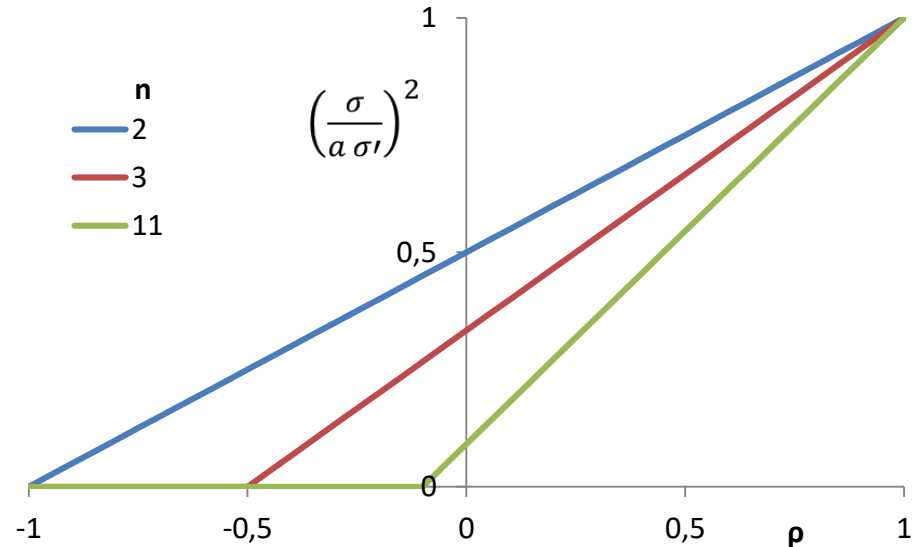
homogene standaardafwijkingen σ en correlatiecoëfficiënten ρ

$$x = \sum a_i x_i$$

$$\sigma^2 = \sum a_i a_j \rho_{ij} \sigma_i \sigma_j$$

$$\left(\frac{\sigma}{a \sigma'}\right)^2 = n(1 + (n-1)\rho)$$

$$\rho \geq \frac{-1}{n-1}$$



- dus $\rho = 0$ en $\rho = 1$ zijn acceptabele extremen
- σ en $\mathbf{P}$ homogeen zijn weinig realistische aannamen

quasi-homogene σ en ρ

blok- en banddiagonaalmatrices

- lege cellen bevatten consistente ρ
- fysieke structuur vervangen door correlatiestructuur

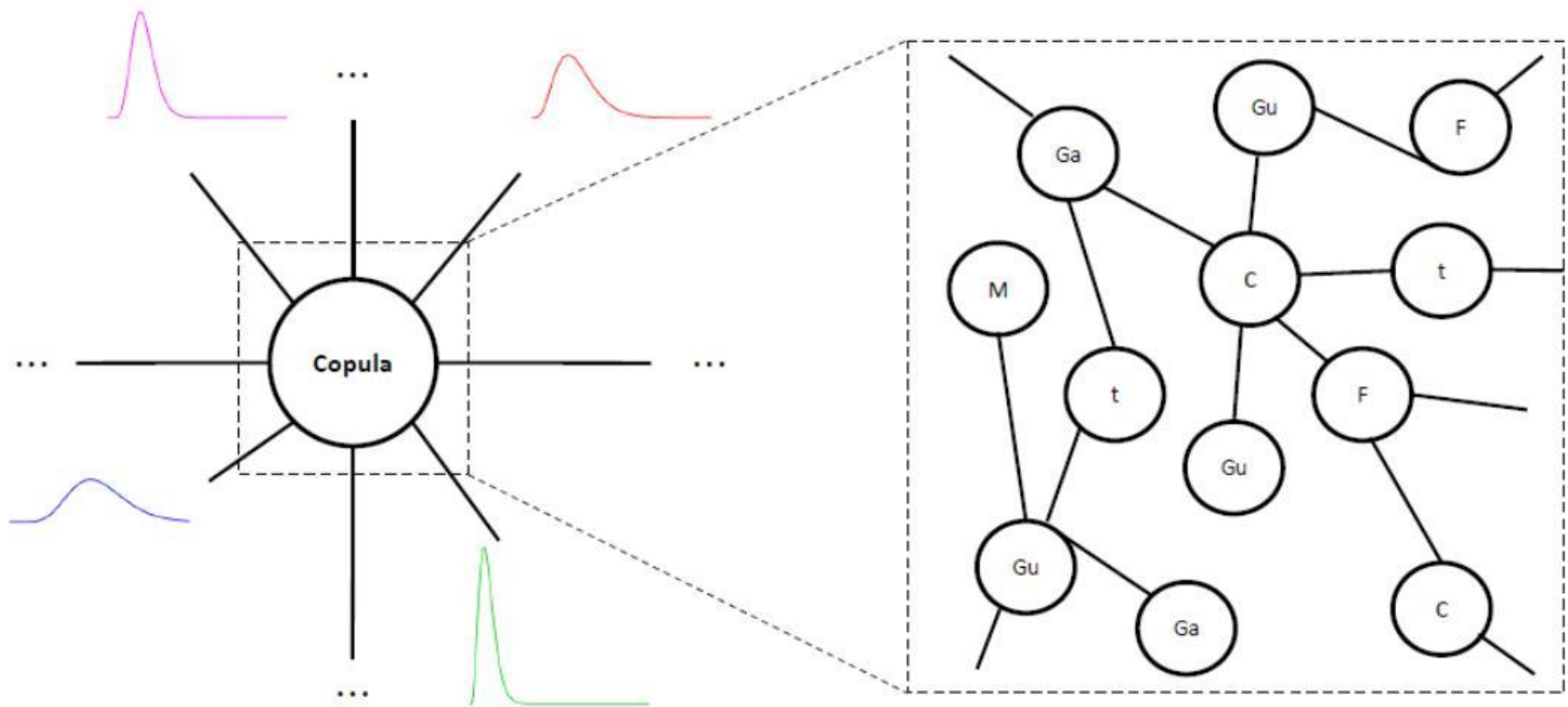
1	a	a	a			
a	1	a	a			
a	a	1	a			
a	a	a	1	b	b	
			b	1	b	
			b	b	1	c
					c	1

1	ρ					
ρ	1	ρ				
	ρ	1	ρ			
		ρ	1	ρ		
			ρ	1	ρ	
				ρ	1	ρ
					ρ	1

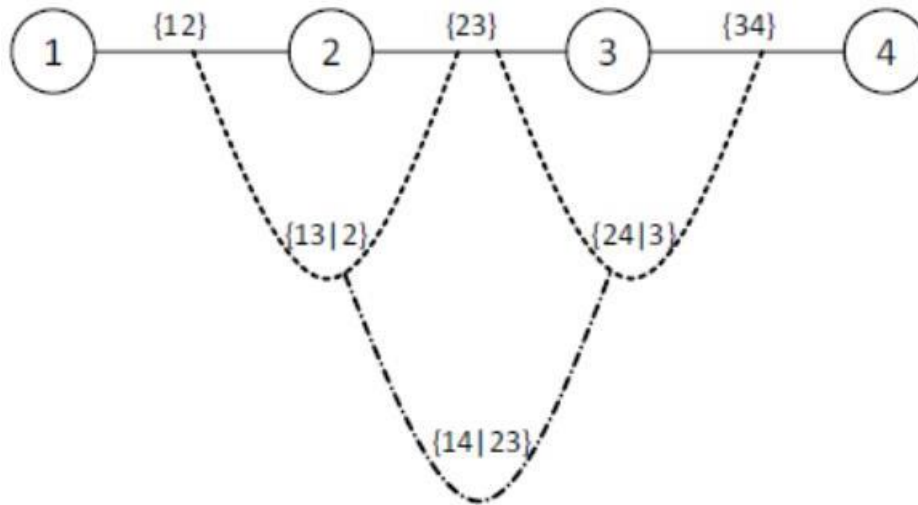
netwerk van variabelen, casusstructuur (Haff 2016)

Complete multivariate distribution

Pair-copula construction



conditionele correlatie, D-vine (Joe 1996)

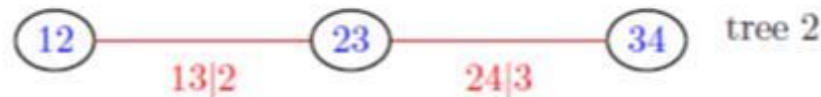


$$\begin{pmatrix}
 1 & \rho_{12} & \rho_{13} & \rho_{14} & \rho_{15} \\
 \rho_{12} & 1 & \rho_{23;1} & \rho_{24;1} & \rho_{25;1} \\
 \rho_{31} & \rho_{32} & 1 & \rho_{34;12} & \rho_{35;12} \\
 \rho_{41} & \rho_{42} & \rho_{43} & 1 & \rho_{45;123} \\
 \rho_{51} & \rho_{52} & \rho_{53} & \rho_{54} & 1
 \end{pmatrix}$$

D-vine

(drawable-vine), Kraemer, Schepsmeier (2011)

$$f_{1234} = \underbrace{f_1 \cdot f_2 \cdot f_3 \cdot f_4}_{\text{nodes in } T_1} \cdot \underbrace{c_{12} \cdot c_{23} \cdot c_{34}}_{\substack{\text{edges in } T_1 \\ \text{nodes in } T_2}} \cdot \underbrace{c_{13|2} \cdot c_{24|3}}_{\substack{\text{edges in } T_2 \\ \text{nodes in } T_3}} \cdot \underbrace{c_{14|23}}_{\text{edge in } T_3}$$



overblijvend probleem: hoe belangrijk zijn correlaties?

- numeriek $n < \left(\frac{\sigma}{a \sigma'}\right)^2 < n^2$
- determinisme – probabilisme
 - ... It is pretty easy to be certain.
One has only to be sufficiently vague. ...
[Peirce (1839 – 1914), Collected Papers]
- empirisch, vaak onderscheiden in
 - extremen
 - rampen (bijv. Springtij 1953, Katherina 2005 / tsunami 2004, 2008 / brug Genua 2018 / grote projecten; meestal rampen, maar soms successen)
 - financiële modellen
 - normalen
 - hier geen voorbeelden, enkele algemene stellingen
 - Murphy (Finagle, Sod): anything that can go wrong will go wrong (at the worst possible time)
 - Saaty: wisdom of the crowd: (uitmiddelen mee- en tegenvallers, onbewuste verbanden)



conclusies en stelling

- bereikte doelen
 - PEM (point estimate model), simultane verdeling
 - consistentie (positieve correlatiematrix)
 - casusstructuur vervangen door correlatiestructuur
 - triangulatie (vine) -> successievelijke voorwaardelijke correlatiecoëfficiënten
- toekomst: belang van
 - ρ t.o.v. de 2 andere factoren van een risicobeschouwing:
 - casusstructurering
 - marginale verdelingen (min-modus-max, 'driehoeksverdeling')
 - probabilistiek en risico i.h.a.
- modellering van kosten en tijd is nogal primitief t.o.v. constructie-modellen (daar tenminste expliciete partiele factoren met soms impliciete correlatie, 'veiligheidsfactoren')

... er zijn veel mislukte projecten en maar zo nu en dan een brug ...