Tools for Elicitation of Continuous Probability Distributions for Bayesian Networks

Marek J. Druzdzel

BayesFusion, LLC

marek@bayesfusion.com
http://www.bayesfusion.com/

Overview

- Hybrid Bayesian networks
- Metalog distributions
- Expressions producing probability distributions
- Concluding remarks

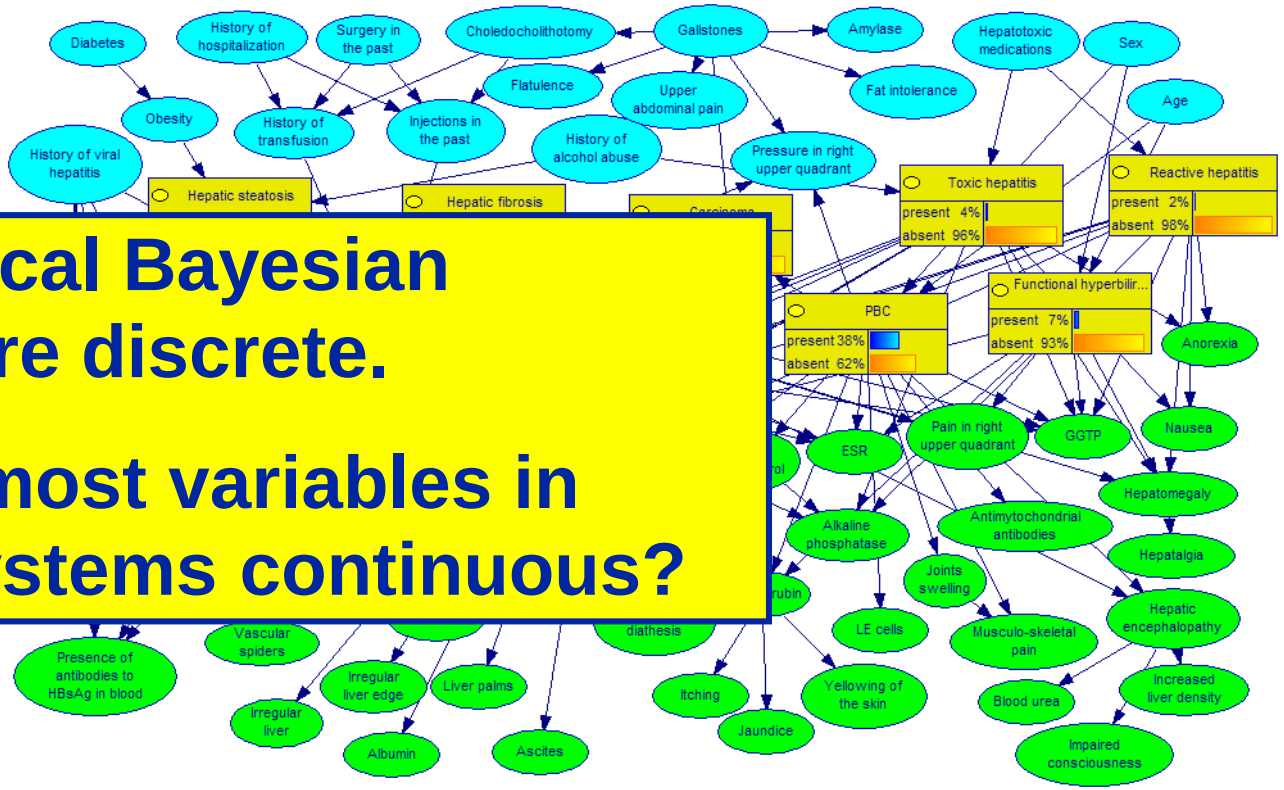
A part of the presentation will be an interactive demo of GeNIe and helpful web tools

Hybrid Bayesian Networks

Bayesian networks: Hepar II Model

The graphical part of a Bayesian network is a representation of causal relations among the model variables

Bayesian networks lead to enormous savings in representation of joint probability distributions



Most practical Bayesian networks are discrete.

But aren't most variables in physical systems continuous?

[Oniśko et al.] 70 variables, 123 arcs, 2,415 independences, 2,139 numerical parameters (instead of over $2^{70} \approx 10^{21}$!)

Equation-based systems and graphical models

classsize = (*nstud* * *cload*) / (*nfac* * *tload*)

facsal = (*oinc* + *tuition* * *nstud*) / (*nfac* * (1 + *overh*))

stratio = *nstud* / *nfac*

← Core equations

cload = 15

tload = 6

nstud = 22102

nfac = 3006

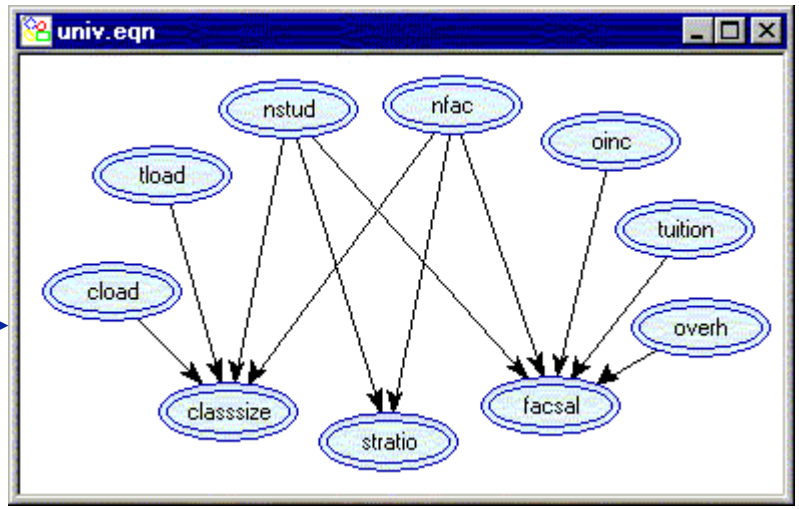
oinc = 30000000

tuition = 12000

overh = 0.48

← Equations for exogenous variables

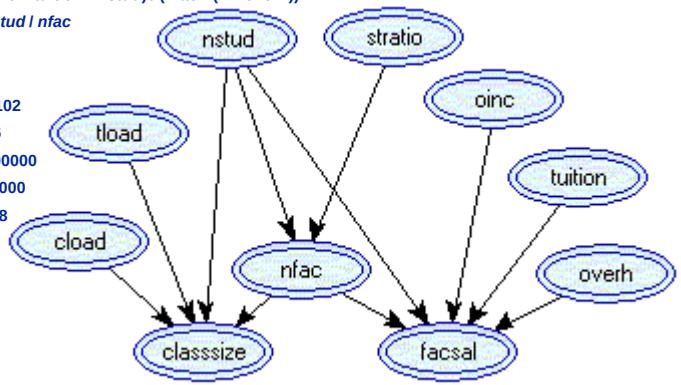
Together they determine the structure of the model →

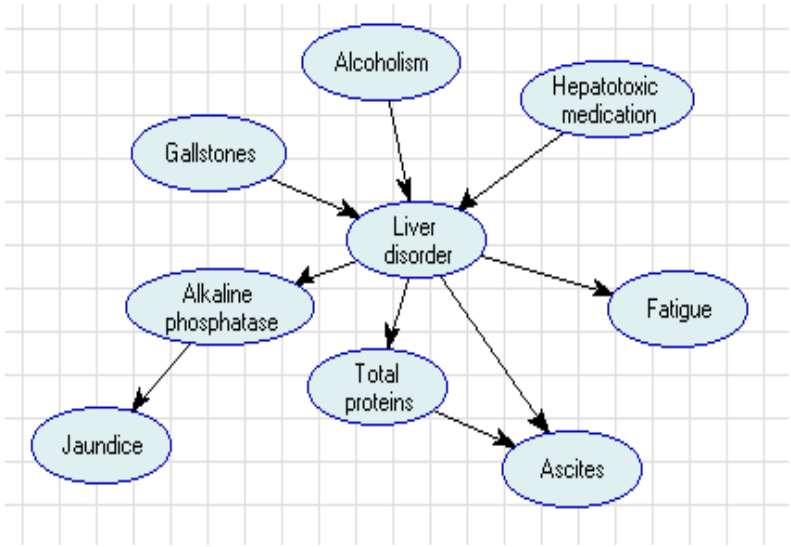


Explication of the asymmetries due to Herb Simon (early 1950s)

Family of directed graphs (a bigger picture)

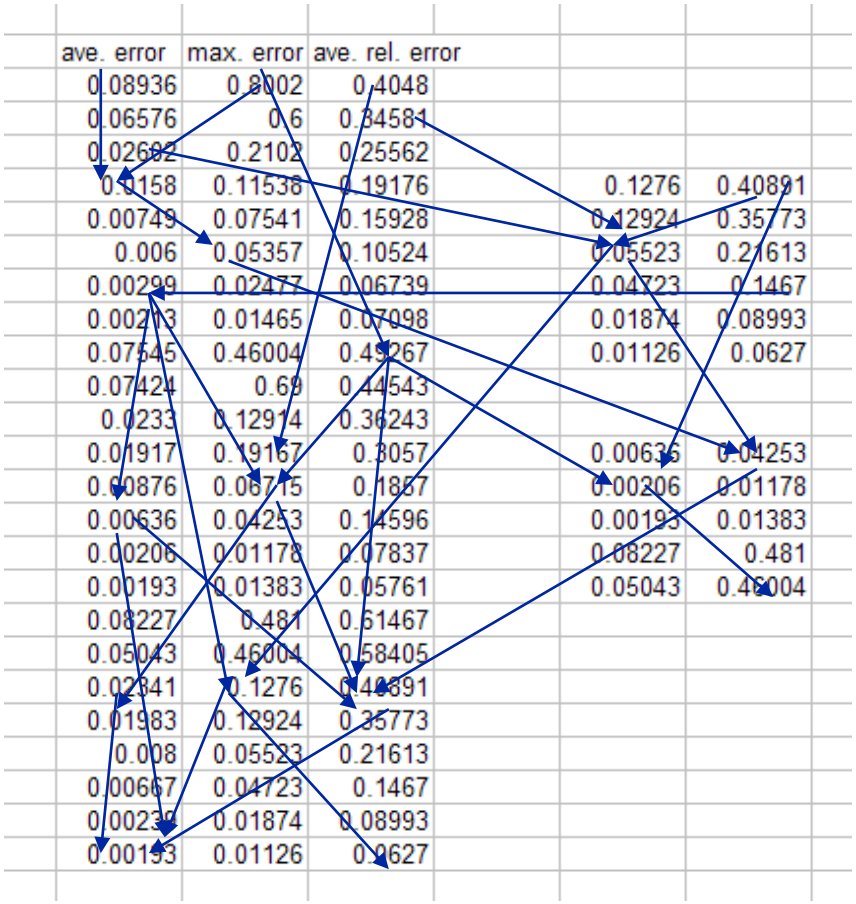
classsize = (*nstud* * *cload*) / (*nfac* * *tload*)
facsal = (*oinc* + *tuition* * *nstud*) / (*nfac* * (1 + *overh*))
stratio = *nstud* / *nfac*
cload = 15
tload = 6
nstud = 22102
nfac = 3006
oinc = 30000000
tuition = 12000
overh = 0.48





Both, systems of equations and joint probability distributions can be pictured by acyclic directed graphs.

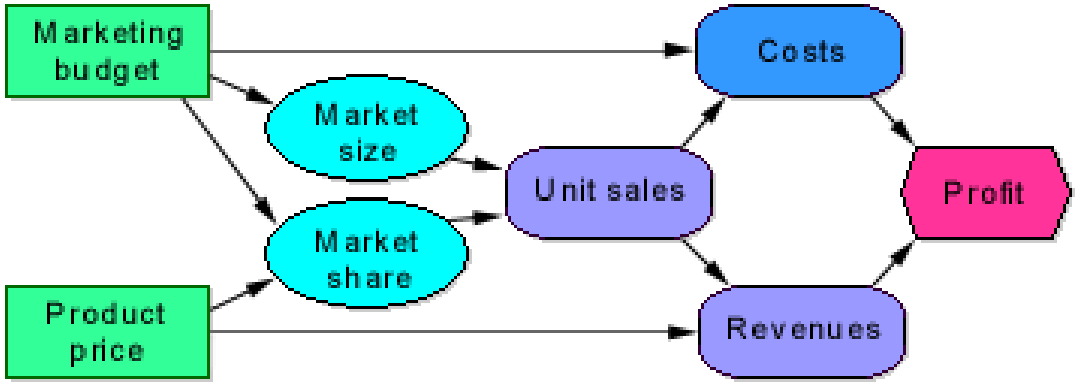
Spreadsheet models



ave. error	max. error	ave. rel. error		
0.08936	0.8002	0.4048		
0.06576	0.6	0.34581		
0.02682	0.2102	0.25562		
0.0158	0.11538	0.19176	0.1276	0.40891
0.00749	0.07541	0.15928	0.12924	0.35773
0.006	0.05357	0.10524	0.05523	0.21613
0.00299	0.02477	0.06739	0.04723	0.1467
0.00213	0.01465	0.07098	0.01874	0.08993
0.07545	0.46004	0.49267	0.01126	0.0627
0.07424	0.69	0.44543		
0.0233	0.12914	0.36243		
0.01917	0.19167	0.3057	0.00636	0.04253
0.00876	0.06715	0.1857	0.00206	0.01178
0.00636	0.04253	0.14596	0.00193	0.01383
0.00206	0.01178	0.07837	0.08227	0.481
0.00193	0.01383	0.05761	0.05043	0.46004
0.08227	0.481	0.61467		
0.05043	0.46004	0.58405		
0.02341	0.1276	0.40891		
0.01983	0.12924	0.35773		
0.008	0.05523	0.21613		
0.00667	0.04723	0.1467		
0.00239	0.01874	0.08993		
0.00193	0.01126	0.0627		

- They could also be viewed as graphs
- Graphs would show causal dependences among cells (variables)
- Of course, for any practical spreadsheet, we would essentially get a spaghetti of connections ☺
- Systems of simultaneous equations and spreadsheet models are not the best we can do
- Directed graphs seem to be better as a user interface!

Visual spreadsheets



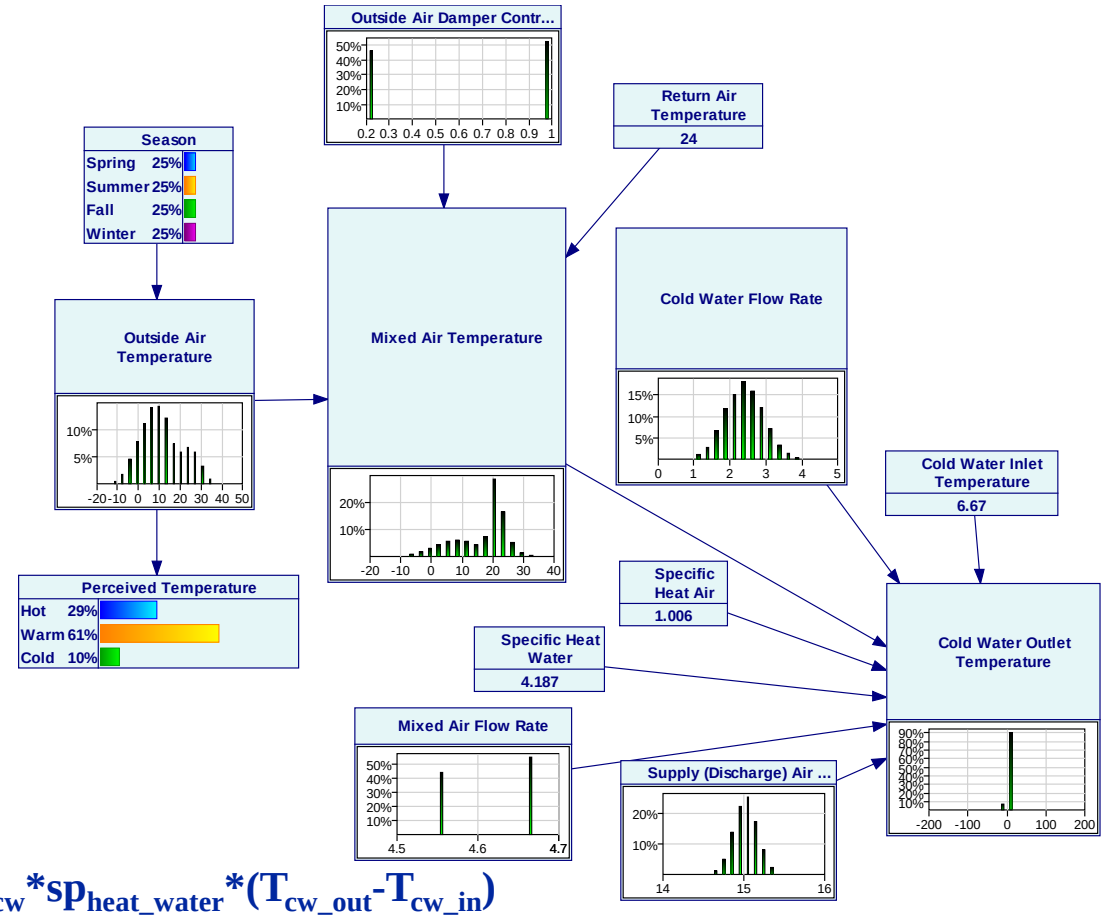
- **Fix almost everything that has been wrong with spreadsheets**
- **Great, but I believe that they could still be improved upon!**

My favorite is Analytica (http://www.lumina.com/)

A simultaneous structural equation-based model cam be turned into a Bayesian network

A model of heating and cooling of buildings.

Two core equations, continuous variables/distributions.



Equations relating temperatures before and after the damper:

$$T_{ma} = T_{oa}*u_d + T_{ra}*(1-u_d)$$

If there is only cooling (u_{hc}=0)

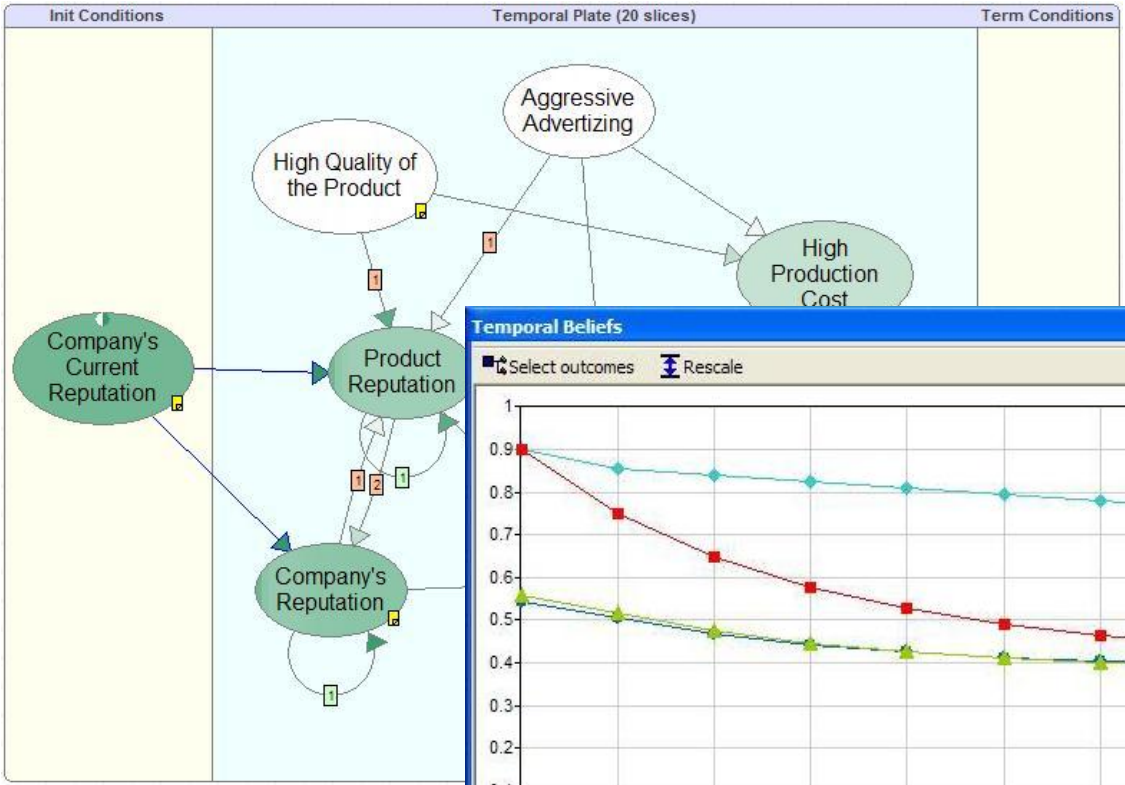
$$m_{flow\_ma}*sp_{heat\_air}*(T_{sa}-T_{ma}) = mdot_{cw}*sp_{heat\_water}*(T_{cw\_out}-T_{cw\_in})$$

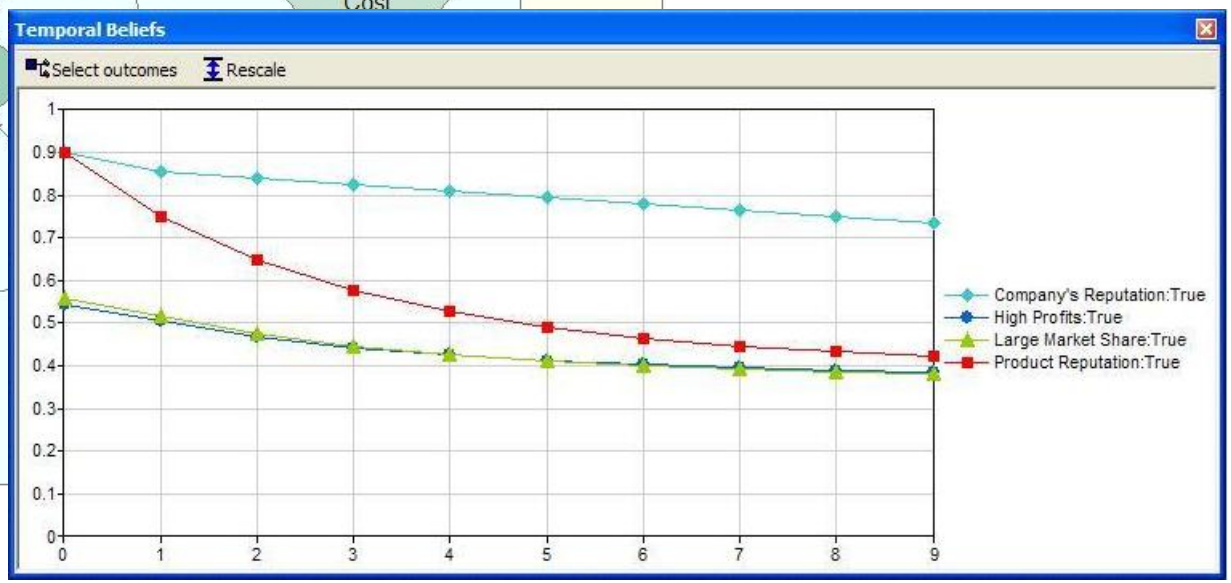
and if there is only heating (u_{cc}=0)

$$m_{flow\_ma}*sp_{heat\_air}*(T_{sa}-T_{ma}) = mdot_{hw}*sp_{heat\_water}*(T_{hw\_out}-T_{hw\_in})$$

Temporal reasoning: Dynamic Bayesian networks

Dynamic Bayesian networks allow for tracking development of a system over time and support decision making in complex environments, where not only the final effect counts but also the system's trajectory.



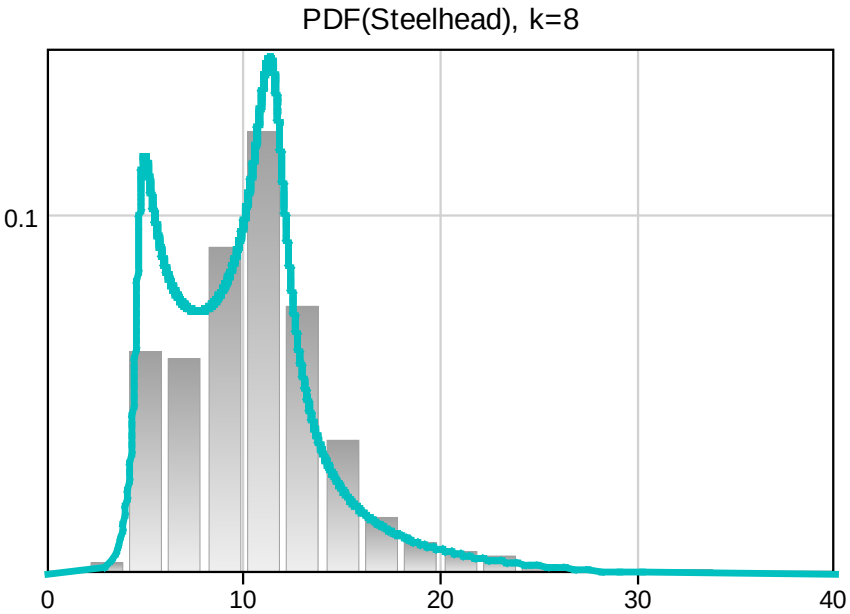


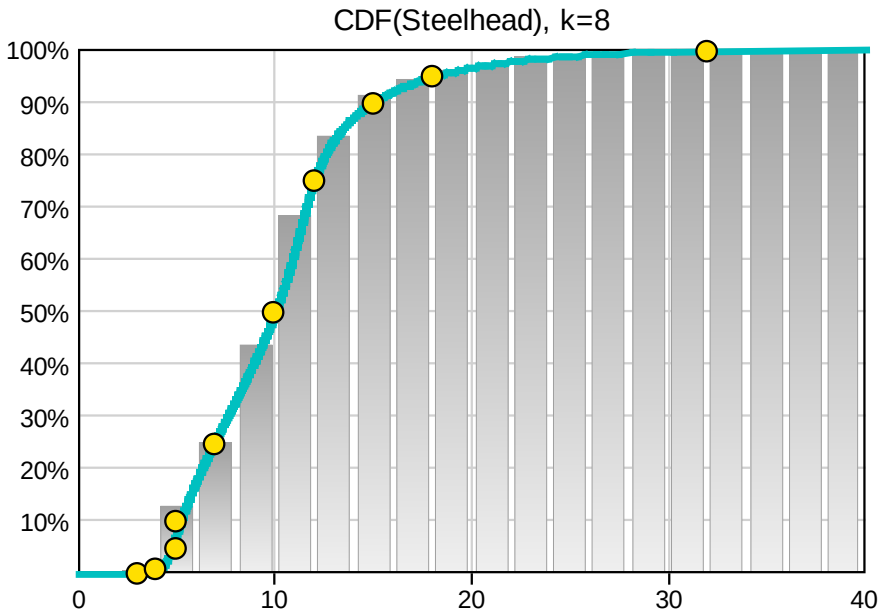
Inspired by systems of differential equations

(the ground work for this was laid by Iwasaki & Simon in early 1990s)

Metalog Distributions

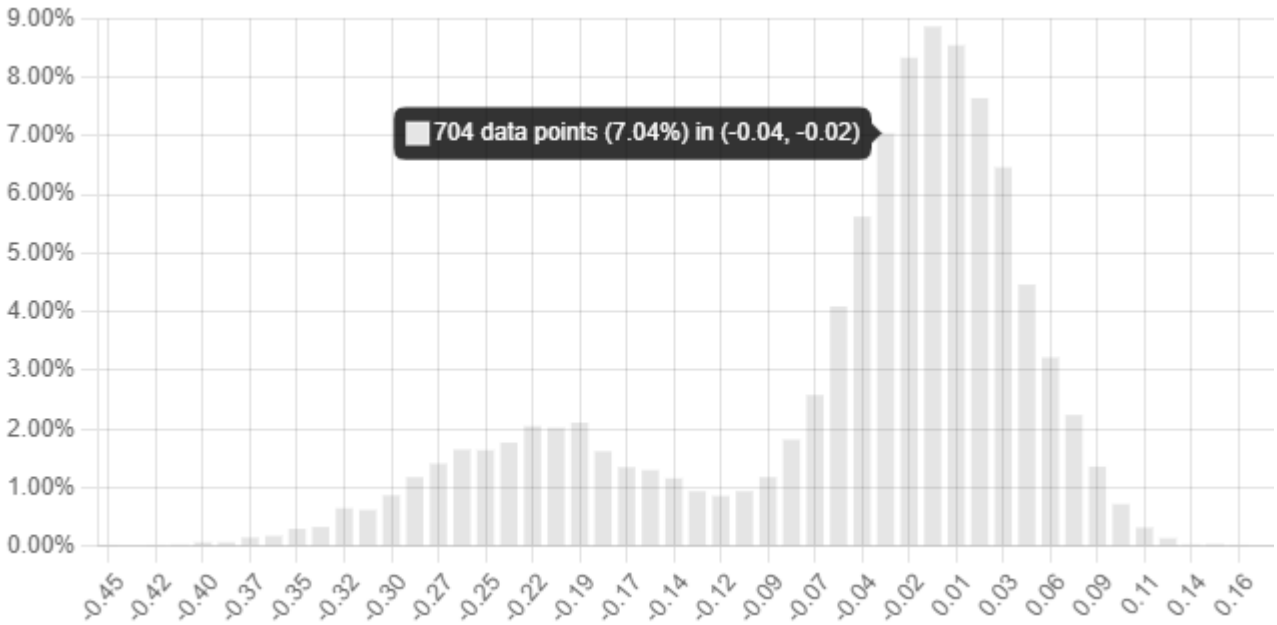
Metalog distributions





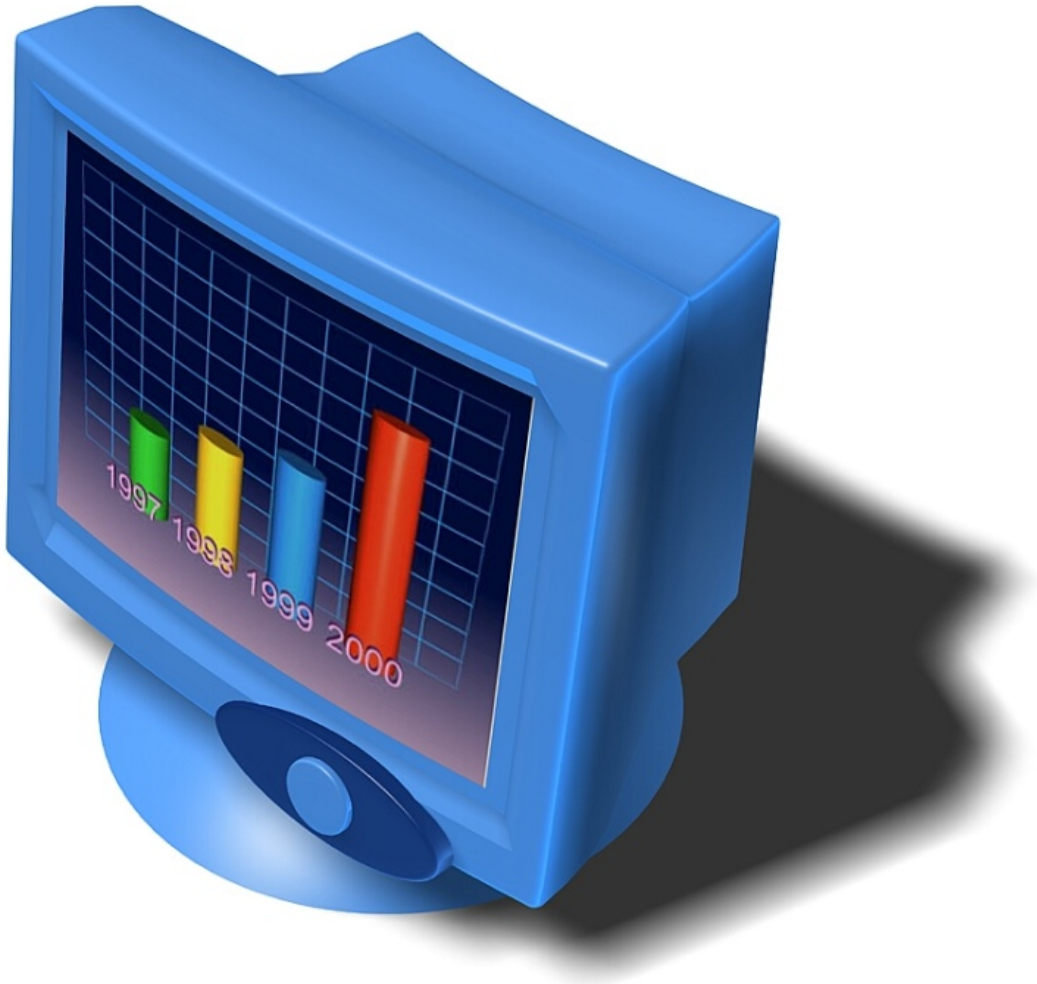
Expressions Producing Continuous Probability Distributions

Visualization of continuous probability distributions



Log10(Sqrt(If(Uniform(0,1)<1/4,Lognormal(-1,0.3),Lognormal(0, 0.2))))

The rest



Concluding remarks

- Continuous variables and distributions are often more natural than their discrete approximations
- The link between systems of simultaneous structural equations and Bayesian networks is often unknown or misunderstood
- Metalog distributions are worth looking at
- *Probability Distribution Visualizer* is a great exploratory tool

You can play with the ideas presented in this talk in GeNIe but also at https://metalog.bayesfusion.com/ and https://prob.bayesfusion.com/

Thank you for your attention!

- http://www.bayesfusion.com/
- bayesfusion
- bayesfusion
- bayesfusion
- BayesFusion
- bayesfusion-llc
- bayesfusion

For technical inquiries please visit http://support.bayesfusion.com/forum/