The Elicitation of Stories

Jim Q. Smith

University of Warwick & Alan Turing Institute

Subjective Bayes Sept 2016

Stories of Unfoldings & Chain Event Graphs

- Because expressible in natural language - **structure easier to faithfully elicit than quantities**.
- Expert judgments often structured within a **story**: when this so - good to elicit this first.
- By embellishing an event tree with colours - changing its topology into a **chain event graph (CEG)** - can directly **express a story formally**.
- A **CEG generalises a discrete BN**. Nevertheless shares with BN nearly all of its desirable properties.
- Can always directly hang elicited probabilities on the CEG & **perform Bayesian inference on it directly**.
- CEGs already provenly **useful in many domains** - Forensic Science, biology, radicalisation processes, public health, ... see e.g. Collazo & Smith(2016) Barclay et al (2013,14) & Collazo et al (2016).
- Here illustrate representational power of a **CEG** & how to use it as a tool **in subjective Bayesian elicitation & inference**.

What I will do

- Discuss **probability trees**
- **Illustrate** how to elicit how things might happen & represent as a CEG: with 2 examples from forensic science & public health.
- Demonstrate how this **unquantified structure** used to encourage client to appraise implications of her statements & adjust these if necessary to make description **requisite**.
- Show that the CEG is a natural structure for expressing **causal conjectures**.
- Show how various tree model hypotheses stand up to data analysis & linking this to **subjective Bayesian Model Selection**.
- Review some recent work to illustrate how the ideas extend to **infinite trees** and semi - Markov processes.

An example of an activity level forensic inference

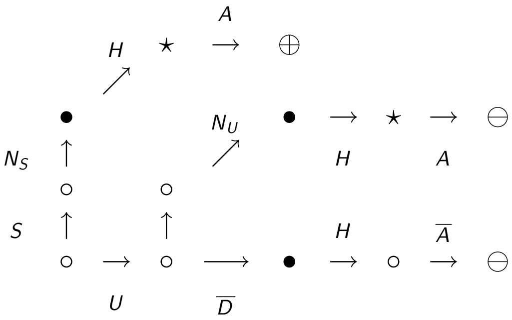
- **Woman** V wearing a recently washed dressing gown attacked by Y at her home at night, assaulted & raped.
- One hair found on V 's dressing gown not her own. All agree DNA matched **suspect** S's: match discovered after search of national database. Other evidence points to undisputed fact that this hair donated during assault.
- V & S **were strangers** & no reason to meet or for S to be at house legitimately. So V could not have donated S's hair herself.
- S claims not to be Y nor to be in a nearby area at time of assault & that hair from **some other unknown person** U.

Prosecution H_p : S assaulted V

Defense H_d : U assaulted V

Non-zeroed edges of event tree of case + Notation

$N_S (N_U) \triangleq S\ (U)$ **nearby** when crime took place
$H \triangleq$ **one hair** from Y retrieved from V
$A \triangleq$ hair retrieved hair belongs to **assailant** .
$D \triangleq$ **DNA** of S & U match



Chain Event Graphs in General

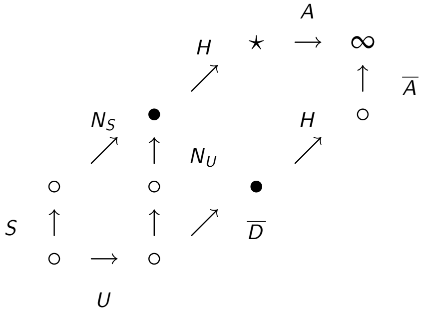
- Derived from probability trees but often **topologically much simpler**.
- Like a tree embed collections of hypotheses about **how things might have happened**.
- Like a **tree paths** represent fully structure of **sample space**.
- Unlike a tree but like a BN **able to express many hypothesised independences** within the story. These can be read from the **cuts** in the graph Smith& Anderson (08) Collazo et al (16)
- Like a BN **full propagation** algorithms available for fast probabilistic reasoning even in very complex scenarios.
- Like BNs provide a **framework for conjugate inference** & model selection.

Chain Event Graphs for Forensic Science

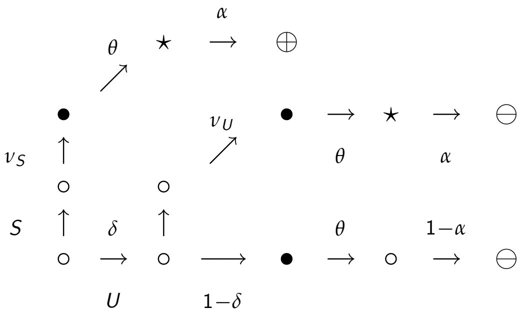
- Even in simple forensic cases events that matter (& so the relevant rvs) to defense are different to those of prosecution. e.g. here existence of U sharing $S's$ dna only comes into defence propositions. So **asymmetric.**
- Such **asymmetries multiply with complexities** of case or with composite propositions.
- This asymmetry is **very difficult to capture using a BN** without creating many zero prob (& often nonsense) events. CEG captures this directly
- Unlike tree, expresses **conditional independences** (from identified edge probs) within its topology & colouring!

Non zeroed edges of CEG after evidence

$N_S(N_U) \triangleq S\ (U)$ nearby when crime took place $P(N_x) \triangleq \nu_x$
$H \triangleq$ one hair from Y retrieved from V - $P(H) \triangleq \theta$
$A \triangleq$ hair retrieved hair belonging to assailant $P(H) \triangleq \alpha$.
$D \triangleq$ DNA of S & U match - $P(D) \triangleq \delta$



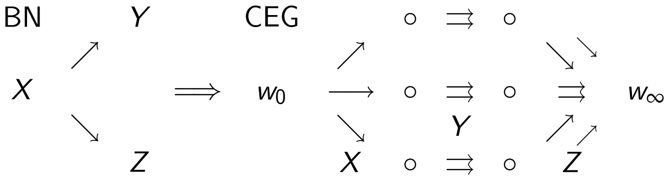
The likelihood ratio of the case



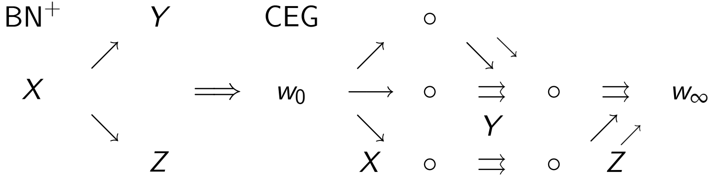
$$LR = \frac{P(\oplus)}{P(\ominus)} = \frac{\nu_S \theta \alpha}{\delta \nu_U \theta \alpha + (1-\delta)\theta(1-\alpha)} = \frac{\nu_S \alpha}{\delta \nu_U \alpha + (1-\delta)(1-\alpha)}$$

Aside: CEG which extends a BN

.



but context specific BN^+ fits much better



(distribution of Z same whether or not X takes medium or large value)

Read CI from CEG?(Smith & Anderson,08, Thwaites & Smith,15)

Theorem

If the random variables $X_1, X_2, \ldots, X_n$ with known sample spaces are fully expressed as a BN, G, or as a context specific BN G, and you know its CEG, C, then the random variables $X_1, X_2, \ldots, X_n$ and all their conditional independence structure together with their sample spaces can be retrieved from C.

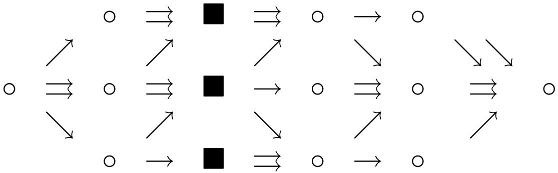
Theorem

$Downstream \amalg Upstream|\ w-Cut$

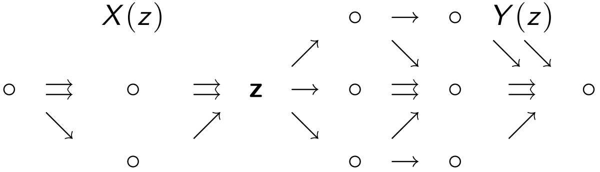
Theorem

$Children \amalg Upstream|u-Cut$

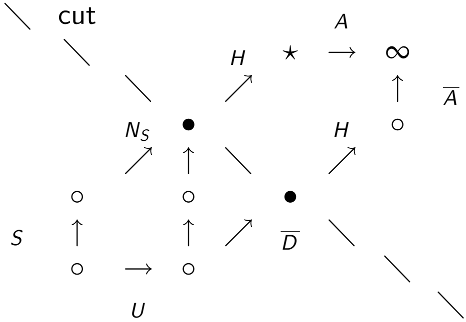
Example of a CEG with Cuts



Downstream $Y(z)$ independent of upstream $X(z)$ given cut $Z = z$.Cuts need not be orthogonal. So can construct dependence through functional relationships.



Example of a cut in our CEG



Corollary of Thm. in Smith & Anderson (08) reads from CEG "innocence or guilt of our suspect does not depend on θ." Note in LR θ cancels out

$$\frac{P(\oplus)}{P(\ominus)} = \frac{\nu_S \theta \alpha}{\delta \nu_U \theta \alpha + (1-\delta)\theta(1-\alpha)} = \frac{\nu_S \alpha}{\delta \nu_U \alpha + (1-\delta)(1-\alpha)}$$

So indeed the case!

Causal Bayesian Networks

- Recall that for causal BNs
  - Variables not downstream of X, a manipulated node, are unaffected by the manipulation.
  - X is set to the manipulated value $\hat{x}$ with probability 1.
  - Effect on downstream variables is identical to ordinary conditioning.
- But many manipulations don't follow these rules, e.g. "Whenever a unit is in set A of positions, take it to another position B".

Causal CEGs

- Can be implemented on a CEG by making paths through a position w pass along a designated edge to a designated position w' (retain all other floret distributions).
- Similarly to BNs:
  - Probs of edges not after w unchanged.
  - An edge from w to w' forces w' after w.
  - Downstream probabilities after w' unchanged.
- Graph of CEG tells us when can find Bayes estimate of effect of a manipulation when unmanipulated system only partially observed
  - Generalizations of Pearl's Backdoor Theorem now proven Thwaites et al(2010), Thwaites (2012).

So only **qualitative structure of CEG** needed to answer such questions!!!

Drawing experimental and sample evidence into CEG's

- Likelihood separates! so class of regular CEG's admits simple conjugate learning.
- For example likelihood under complete random sampling given by

$$
\begin{aligned}
l(\boldsymbol{\pi}) &= \prod_{u \in U} l_u(\boldsymbol{\pi}_u) \\
l_u(\boldsymbol{\pi}_u) &= \prod_{i \in u} \pi_{i,u}^{x(i,u)}
\end{aligned}
$$

  where $x(i,u)$ $\#$ units entering stage u & proceeding along edge labelled (i,u), $\sum_i \pi_{u,i} = 1$ in sample.
- From Bayesian perspective e.g. independent Dirichlet priors $D(\boldsymbol{\beta}(u))$ on the vectors $\boldsymbol{\pi}_u$ leads to independent Dirichlet $D(\boldsymbol{\beta}^*(u))$ posteriors where

$$\beta^*(i,u) = \beta(i,u) + x(i,u)$$

Conjugate Bayesian Inference on CEG's

- Prior stage floret independence a generalisation of local & global independence in BNs. Just as in Geiger & Heckerman(1997), floret independence, + appropriate Markov equivalence characterises product Dirichlet prior (see Freeman and Smith, 2011a).
- Under characterisation only a small no. of prior parameters over whole model class: so domain judgements can be specified through one & extended to many.
- Just like for BNs, data from undesigned experiments or poorly randomised surveys or using non - ancestral sampling of a CEG data destroys conjugacy, but inference is no more difficult than for a BN.

Learning the topology of a CEG

.

- Using appropriate priors on model space & modular parameter priors over CEGs, log marginal likelihood score of complete observational data, experimental data or good surveys *linear* in CEG stage components.
- Explicitly for $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_k)$, let $s(\boldsymbol{\alpha}) = \log \Gamma(\sum_{i=1}^{k} \alpha_i)$ and $t(\boldsymbol{\alpha}) = \sum_{i=1}^{k} \log \Gamma(\alpha_i)$

$$\begin{aligned} \Psi(C) &= \log p(C) = \sum_{u \in C} \Psi_{u(c)} \\ \Psi_{u(c)} &= \sum s(\alpha(i,u)) - s(\alpha^*(i,u)) + t^*(\alpha(i,u)) - t(\alpha(i,u)) \end{aligned}$$

- e.g. MAP model selection using AHC , Dynamic Prog.,Integer Prog, simple & fast over vast space of CEG's (see Cowell & Smith,2014).

Do CEG's fit better than BN's (Barclay et al, 2012)

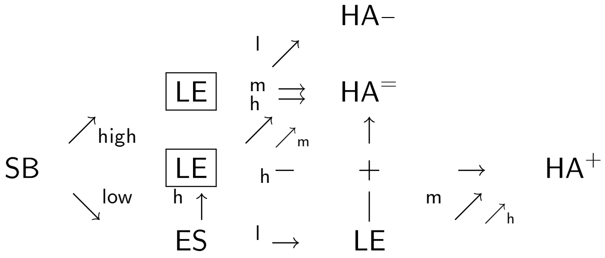
		Social Background		
	$\swarrow$	$\downarrow$	$\searrow$	
Economic Situation	$\longrightarrow$	Family Life Events	$\rightarrow$	Hospital Admissions

.

- Best fit of close competitors: where edges missing from ES$\rightarrow$FLE, & one missing edge into HA.
- Search over all CEGs whose trees consistent with this "causal" order.
- An AHC search allowed us to discover a CEG whose MAP score was 80 times better.

The MAP CEG (omitting sink node)

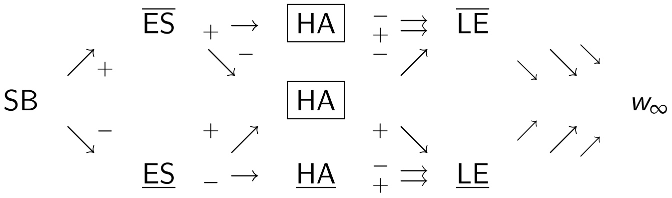
.



- Econ. Sit. not "cause" of life events or hospital admissions for High SB.
- High SB & low LE uniquely "causes" children a favourable HA^-.
- Prob LE for (High SB) & (Low SB +High ES) similar - different HA
- Think of cause in terms of events rather than variables.

Example CHIDS a different CEG

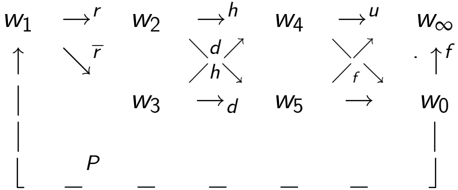
Best model identified through Dynamic Programming allowing changed response variable.



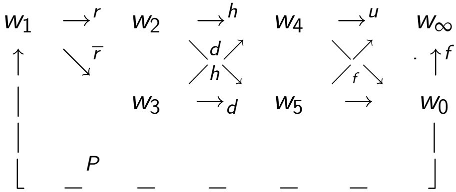
- This model sees *life events as a result of poor child health*.
- Increased incidents of hospital admissions relates only to poverty (2 categories).
- High life events unaffected by Hospital Admissions except that when exactly one of SB or ES is low then poor child health can shift into lower life event category.

Example of a DCEG: rubella cycle

w_0 - she decides to try to get pregnant: edges from positions $\{w_4, w_5\}$.
w_1 - she gets pregnant: edge from position $\{w_0\}$.
w_2 - birth after she caught rubella in the first 3 months of pregnancy: edge from position $\{w_1\}$.
w_3 - normal birth: edge from position $\{w_1\}$.
w_4 - hearing baby: edges from positions $\{w_2, w_3\}$.
w_5 - dead/ deaf baby: edges from positions $\{w_2, w_3\}$.
w_∞ - decides/ unable to further conceive the edges from positions $\{w_0, w_4, w_5\}$



Example of a DCEG: rubella cycle



- DCEG of this type a coloured transition diagram of a semi-Markov process
- w_1 position entered only though $w_0 \Rightarrow$ rubella event no direct impact on future pregnancy.
- To try for more children fn. only on deafness of last child.
- Time here local to each woman. So semi-Markov process draws evidence together across different cases.

Concluding remarks about CEGs

.

- **Trees & CEGs** are a much neglected but **powerful elictation methods** for addressing real elicitation problems.
- **Express & explore hypotheses, synthesise information & evaluate strength of evidence** for & against various hypotheses .
- Whatever you can do for discrete BNs you can also do using **CEGs**
- **CEG software soon on CRAN**. inc. propagation & estimation.

Thank You !!!!!!!!!!!!!

Selected Publications by me

Mazumber, A. & Smith. J.Q.(2016) "Using chain event graphs to address asymmetric evidence in legal reasoning" (in preparation)

Collazo, R.A. & Smith, J.Q.(2016) "A new family of Non-local Priors for Chain Event Graph model selection" Bayesian Analysis (to appear)

Collazo, R.A., Gorgen,C. & Smith, J.Q.(2016) "Chain Event Graphs" Chapman and Hall (to appear)

Barclay, L.M., R. Collazo, Smith, J.Q. Thwaites , P. and Nicholson , A. (2015) "Dynamic Chain Event Graphs" Electronic Journal of Statistics 2015, Vol. 9, 2, 2130-2169.

Cowell, R.G. and Smith, J.Q. (2014) "Causal discovery through MAP selection of stratified chain event graphs" Electronic J of Statistics vol.8, 965 - 997

Barclay, L.M. , Hutton, J.L. and Smith, J.Q.(2013) "Refining a Bayesian Network using a Chain Event Graph" International J. of Approximate Reasoning 54, 1300-1309.

Selected Publications by me

Freeman, G. & Smith, J.Q. (2011a) " Bayesian MAP Selection of Chain Event graphs" J. Multivariate Analysis, 102, 1152 -1165

Thwaites, P. Smith, J.Q. and Riccomagno, E. (2010) "Causal Analysis with Chain Event Graphs" Artificial Intelligence, 174, 889–909

Riccomagno, E.& Smith, J.Q. (2009) "The Geometry of Causal Probability Trees that are Algebraically Constrained" in "Optimal Design & Related Areas in Optimization and Statistics" Eds L. Pronzato & A.Zhigljavsky, Springer 131-152

Thwaites, P., Smith, J.Q. & Cowell, R. (2008) "Propagation using Chain Event Graphs" Uncertainty in Artificial Intelligence, Eds D. McAllester & P. Myllymaki, 546 -553

Smith, J.Q. & Anderson P.E. (2008) "Conditional independence and Chain Event Graphs" Artificial Intelligence, 172, 1, 42 - 68

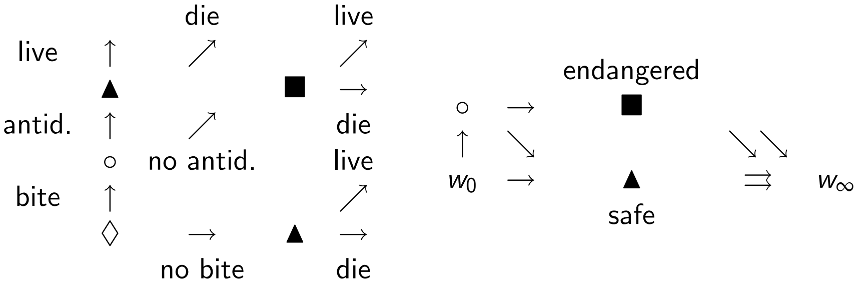
Constructing a CEG

Event tree $\rightarrow$ Staged tree $\rightarrow$ CEG [by positions and stages]

- Start with an event tree as illustrated above.
- Colour the vertices of tree to rep its stages (=staged tree).
- Identify positions (with w_∞ the vertices fo the CEG.
- Construct CEG by inheriting edges in obvious way from tree and attach all leaes to w_∞.

Snake Bite Example: Causal Variables Implicit

$X_1 \sim$ Bitten by snake, $X_2 \sim$ Carry and apply perfect antidote, $X_3 \sim$Die tomorrow..



$X \sim$ not bitten/ bitten but apply antidote, $Y \sim (= X_3)$ live/die, $Z \sim$ safe/endangered.
So from the CEG preferred variables exhibiting the conditional independence can be deduced from graph.