FROM TEXTUAL EVIDENCE TO PROBABILISTIC MODELS

*Kobus Barnard, Clayton Morrison, Adarsh Pyarelal**

IVILab Colloquium, January 19, 2018

Part II

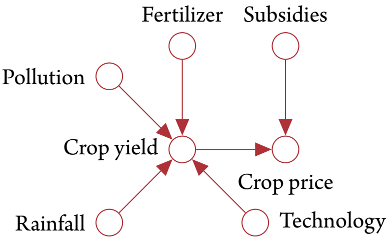
World Modelers

Definition of food insecurity

'Limited or uncertain availability of nutritionally adequate and safe foods or limited or uncertain ability to to acquire acceptable foods in socially acceptable ways'

- Anderson, S. A., *Core indicators of nutritional state for diffcult-to-sample populations*, The Journal of Nutrition (USA) (1990)

Food insecurity is a complex function of many factors, which can be modeled by a *causal influence network*, like the one on the right.



Building on Big Mechanism

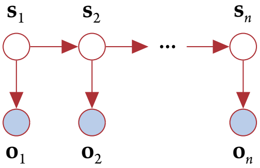
The causal influence network will be assembled from textual evidence, similar to biological pathways.

Making predictions

In addition to understanding the structure of the causal influence network, we would ideally like to use it to make predictions about the factors in the future.

To do this, we use *dynamic Bayes networks*.

Linear Dynamical Systems



$\mathbf{s}_n$:	Latent state at time n	$\mathbf{A}$:	Linear transition model	$\boldsymbol{\Gamma}, \boldsymbol{\Sigma}, \mathbf{V}_0$:	Covariance matrices
$\mathbf{o}_n$:	Observed state at time n	$\mathbf{C}$:	Linear observation model	$\boldsymbol{\mu}_0$:	Mean vector

$$p\left(\mathbf{s}_n | \mathbf{s}_{n-1}\right) = \mathcal{N}(\mathbf{s}_n | \mathbf{A}\mathbf{s}_{n-1}, \boldsymbol{\Gamma}) \qquad \text{Transition probability}$$

$$p\left(\mathbf{o}_n | \mathbf{s}_n\right) = \mathcal{N}(\mathbf{o}_n | \mathbf{C}\mathbf{s}_n, \boldsymbol{\Sigma}) \qquad \text{Emission probability}$$

$$p\left(\mathbf{s}_1\right) = \mathcal{N}(\mathbf{s}_1 | \boldsymbol{\mu}_0, \mathbf{V}_0) \qquad \text{Initial value}$$

Model Parameters

$$\boldsymbol{\theta} = \{\mathbf{A}, \boldsymbol{\Gamma}, \mathbf{C}, \boldsymbol{\Sigma}, \boldsymbol{\mu}_0, \mathbf{V}_0\}$$

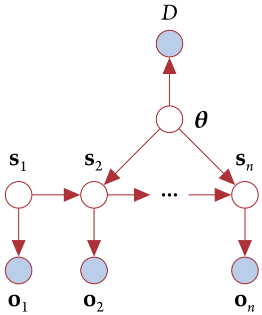
Model Uncertainty

Consider the phrase: *A small increase in X causes a large decrease in Y.*

1. Can we extrapolate from this to talk about the effect of a small *decrease* in *X*?
2. What does the writer mean by 'small' and 'large'? Does their interpretation match that of other writers?
3. What if we have other evidence sentences about *X* and *Y* that differ in
   - Magnitude:
     - Small increase in X causes large increase in Y
     - Tiny increase in X causes huge increase in Y
   - Polarity
     - A small increase in X causes large decrease in Y
     - A small increase in X causes a large increase in Y
4. What if we have no evidence sentences, are there reasonable assumptions?

Idea:
Consider the *model itself* to be a random variable.

Linear Dynamical System with random transition model



Simple causal linear dynamical system

How far can we get with just the following sentence?

A small increase in X causes a large decrease in Y.

Latent states

Our latent states are comprised of the factors and their partial derivatives w.r.t time.

$$\mathbf{s} = \begin{pmatrix} X & \dot{X} & Y & \dot{Y} \end{pmatrix}^T$$
$$\dot{X} = \frac{\partial X}{\partial t}$$
$$\dot{Y} = \frac{\partial Y}{\partial t}$$

Transition model

Our transition model is the following:

$$\mathbf{s}_n = \mathbf{A}\mathbf{s}_{n-1}$$

In the absence of any evidence sentences, the transition matrix $\mathbf{A}$ takes on the following form:

$$\mathbf{A} = \begin{pmatrix} 1 & \Delta t & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \Delta t \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

We will fill in more entries based on textual evidence.

The sentence

A small increase in X causes a large decrease in Y

is telling us about the shape of $\frac{\partial Y}{\partial X}$.

$$
\begin{aligned}
Y_{n+1} &\cong Y_n + \left.\frac{dY}{dt}\right|_n \Delta t \\
&= Y_n + \left(\left.\frac{\partial Y}{\partial t}\right|_n + \left.\frac{\partial Y}{\partial X}\frac{\partial X}{\partial t}\right|_n \right)\Delta t \\
&= Y_n + \left(\dot{Y}_n + \left.\frac{\partial Y}{\partial X}\right|_n \dot{X}_n \right)\Delta t
\end{aligned}
$$

$\dot{Y}$ represents the *intrinsic* tendency of Y to change with time, due to factors that are not explicitly modeled. For now, we set

$$\dot{Y} = 0$$

Thus we have

$$Y_{n+1} = Y_n + \left.\frac{\partial Y}{\partial X}\right|_n \dot{X}_n \Delta t$$

The simplest possibility is that $\frac{\partial Y}{\partial X}$ is independent of $\mathbf{s}_n$.

$$Y_{n+1} = Y_n + \beta_{XY} \dot{X}_n \Delta t$$

Modified transition matrix

Our transition matrix is now

$$\mathbf{A} = \begin{pmatrix} 1 & \Delta t & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & \beta_{XY}\Delta t & 1 & \Delta t \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Assumption: Implication of the dual

If there is only one evidence sentence relating two quantities, it implies its ‘dual’ constructed by reversing the ‘polarity’ of the changes.

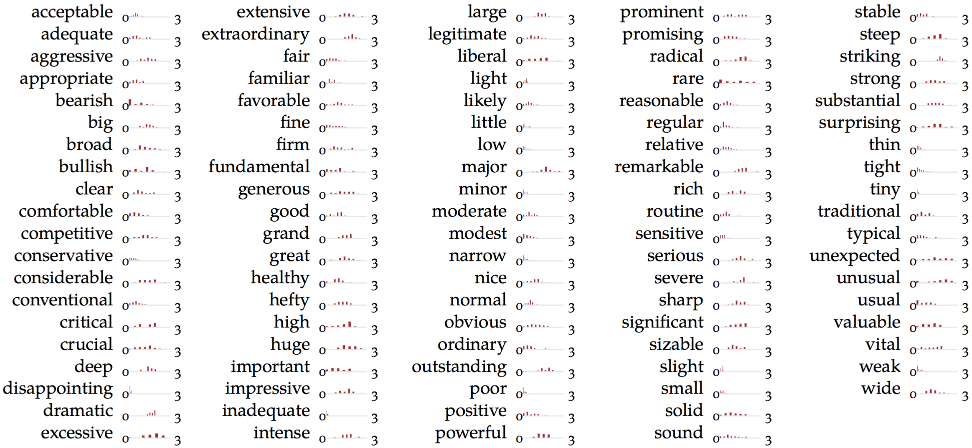
Example:

A small increase in X causes a large decrease in Y.

$$\Updownarrow$$

A small decrease in X causes a large increase in Y.

Grounding gradable adjectives through crowdsourcing

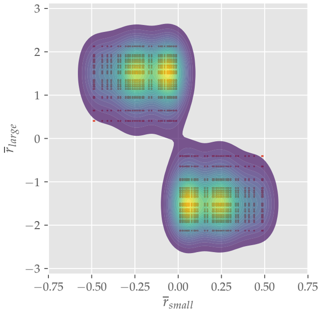


Units: Standard deviations from the mean: $\bar{r} = \frac{r-\mu}{\sigma}$

Joint distribution for two adjectives

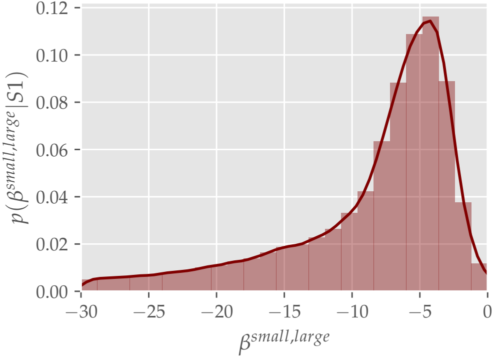
We construct the joint distribution of $\bar{r}$ for the adjectives *small* and *large*, by taking the cartesian product of the responses and then obtaining a Gaussian kernel density estimate. This gives us a distribution over the slopes $\beta^{small,large}$ of lines passing through the origin.

$$\beta_{XY} = \frac{\sigma_X}{\sigma_Y}\beta^{small,large}$$



Empirical distribution of $\beta^{small,large}$

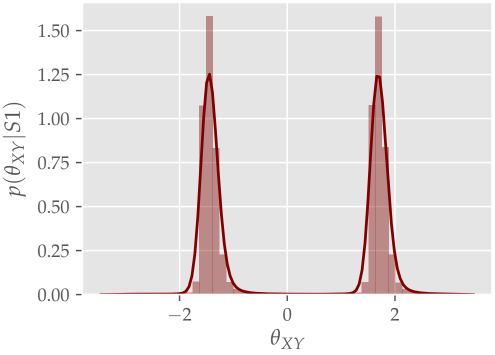
We construct the empirical distribution of $\beta^{small,large}$ from the kernel density estimate.



Reparameterization

We then reparameterize in terms of the polar angle, ensuring compact support for our distribution:

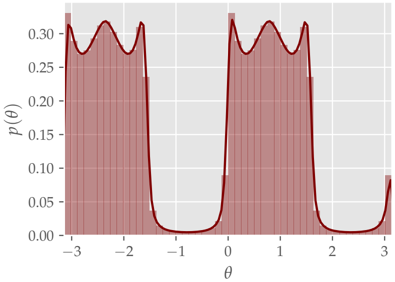
$$\theta_{XY} = \tan^{-1} \beta_{XY}, \theta \in [-\pi, \pi]$$



Prior probability

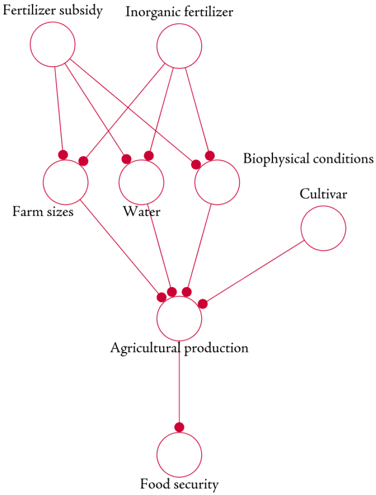
When there are no gradable adjectives to rely on, we use a prior.

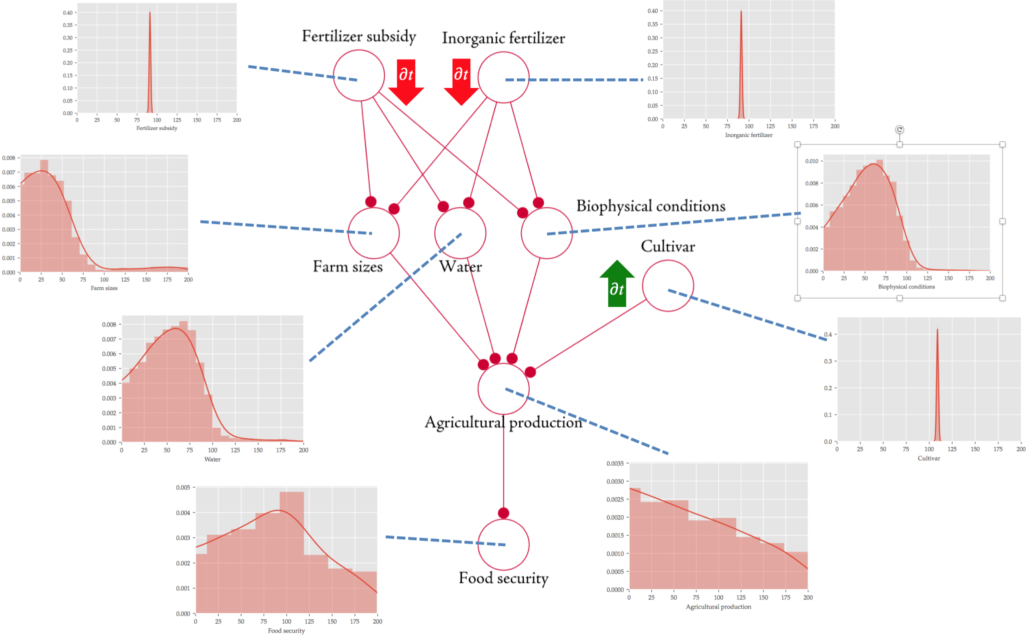
1. For each adjective a:
   1.1 Calculate the kernel density estimator for the collection of responses $\bar{r}$ corresponding to a.
   1.2 Sample n points from the KDE.
2. Collect all the sampled points into a set S. Also construct a set $-S$ containing the negative of the points in S.
3. Construct the set of points $\mathcal{P} = (S \times S) \cup (-S \times -S)$ (for negative correlation, construct $(S \times -S) \cup (S \times -S)$ instead).
4. Take the inverse tangent (restricted to $(-\pi, \pi)$) of the ratio of of their y and x coordinates of each point in $\mathcal{P}$.



A Toy Example

The government promotes improved cultivar to boost agricultural production for ensuring food security. However, the policy to seriously cut down the use of inorganic fertilizer and phase out the fertilizer subsidy results in deteriorating biophysical conditions, low use of inorganic fertilizer, less water, significantly reduced farm sizes which lead to low benefit from the improved cultivar.





Fertilizer subsidy
Inorganic fertilizer
∂t
∂t
Farm sizes
Water
Biophysical conditions
Cultivar
∂t
Agricultural production
Food security
Fertilizer subsidy
Inorganic fertilizer
Farm sizes
Biophysical conditions
Water
Cultivar
Agricultural production
Food security