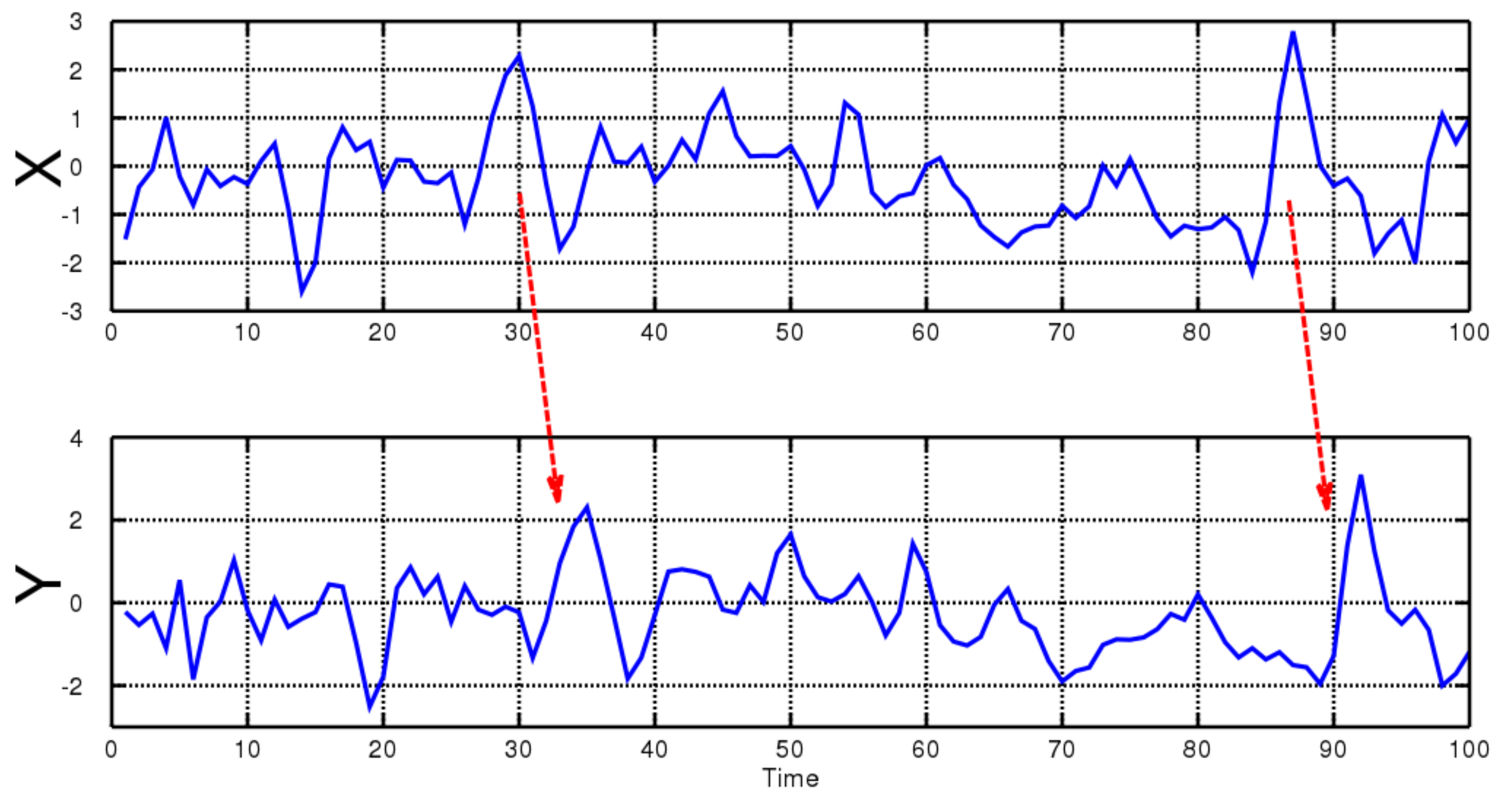


Causality : predictive framework

Bayesian framework for Causality

- Measures of causality : Grainger, Conditional Mutual Information (CMI), Transfer Entropy, Directional Information Transfer, etc.
- We define causality strength as the mutual information normalised by entropy cs = (H(x) - H(x l y)) / H(x) = MI(x,y) / H(x)
- Causality strength is based on finite-length time series it contains estimation errors => pdf of cs
- If new information arises, e.g. new time series, do **NOT** calculate CS for whole time series, but **update** existing knowledge
- Use existing estimate at prior information *p(CS)*
- Calculate CS with uncertainty for new time series, and interpret this as new observation of CS:

$$CS_{obs} = CS_{true} + \epsilon$$

- Use Bayes Theorem to update knowledge

$$p(CS|CS_{obs}) = \frac{P(CS_{obs}|CS)}{p(CS_{obs})} p(CS)$$

Additional causal processes

- We thought the process was

$$x^n = f(x^{n-1}) + \epsilon^{n-1}$$

- The causal connection is quantified in terms of $MI(x^n, x^{n-1})$

- With additional observations of same variable can apply Bayes' theorem to recompute without calculating from scratch

- But new intuition/theory suggests process *y* is also of interest for process *x:*

$$x^n = f(x^{n-1}) + g(y^{n-1}) + \epsilon^{n-1}$$

- For additional process or random variable y, we need to evaluate the influence of both processes (past observations of x and y). Hence we need conditional mutual information given by .

$$CMI(x^n, x^{n-1}|y^{n-1}) \quad CMI(x^n, y^{n-1}|x^{n-1})$$

- Note that the prior for the new connection has to be taken flat.

Problem : Dimension increase of space

$$\text{1D Prior} : p(CS) = p(MI(x^n, x^{n-1}))$$

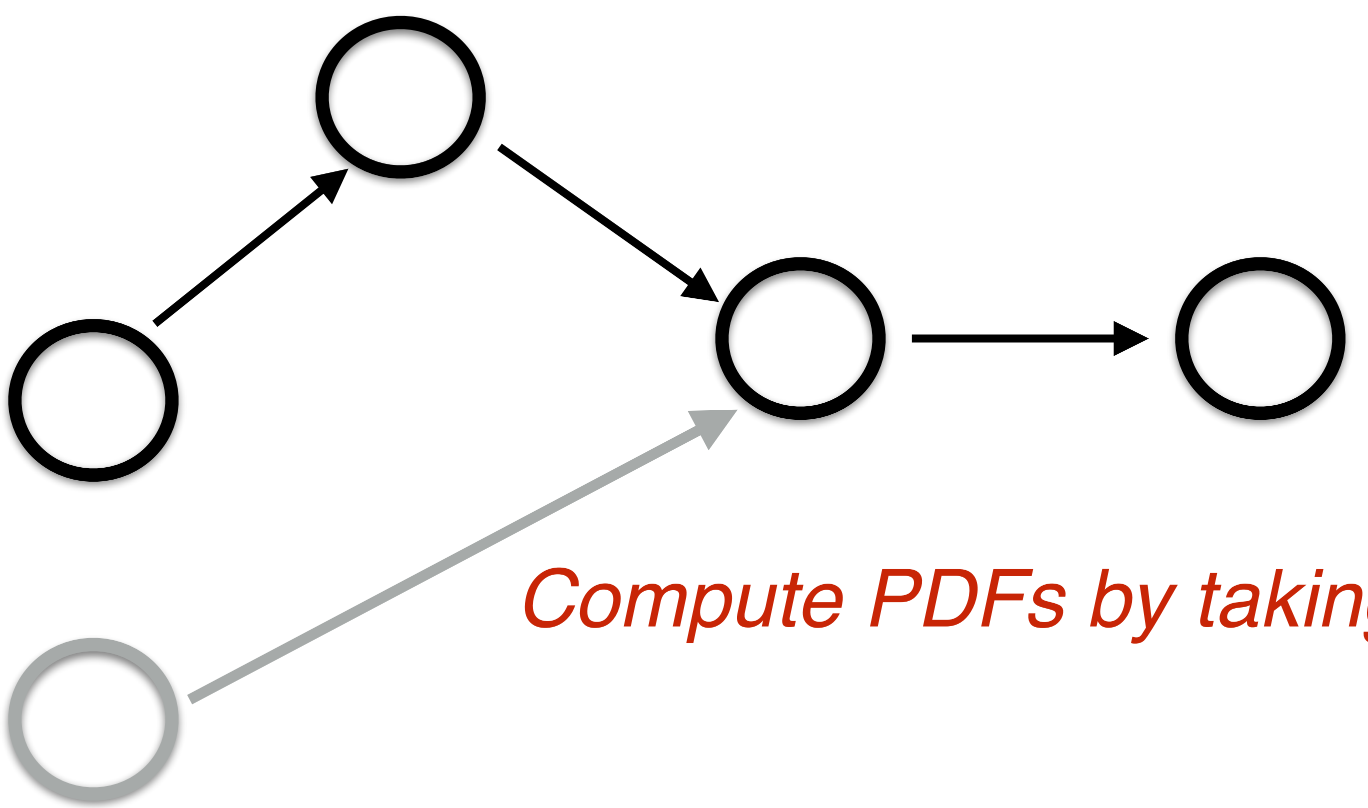
$$\text{2D Likelihood} : p(CS_{obs}|CS)$$

$$\text{where } CS_{obs} = [CMI(x^n, x^{n-1}|y^{n-1}), CMI(x^n, y^{n-1}|n^{n-1})]$$

- Discontinuity in expanding the model (adding dimension)
- PDF in prior encapsulates uncertainty in vector (x^n, x^{n-1})
- PDF in posterior encapsulates uncertainty in vector (x^n, x^{n-1}, y^{n-1})
- *Can't simply use Bayes' theorem !*

Solution….Graphical Representation

- We assert that all processes (random variables) already exist ; we just haven't measured every influence (or causal link)
- Graphical representations provide framework accounting for this space of variables / processes and their relations
- Nodes are random processes ; links contain cs
- Figure - Black nodes are processes with observed ; grey is additional process that appears to have causal influence
- Prior, Likelihood, Posterior distributions on entire network



Compute PDFs by taking different structures

Jon Williamson, 2000 : Foundations for Bayesian Networks

(2007) Causal and Bayesian Networks. In: Bayesian Networks and Decision Graphs. Information Science and Statistics. Springer, New York, NY

Solution….Graphical Representation

- Heckerman shows formal way
- If **m** is model or network structure,

  $p(CS \mid CS_{obs}, \mathbf{m}) = p(CS \mid \mathbf{m})\, p(CS_{obs} \mid CS, \mathbf{m}) \,/\, p(CS_{obs} \mid \mathbf{m})$
  where $p(CS_{obs} \mid \mathbf{m}) = \int p(CS_{obs} \mid CS, \mathbf{m})\, p(CS \mid \mathbf{m})\, d(CS)$

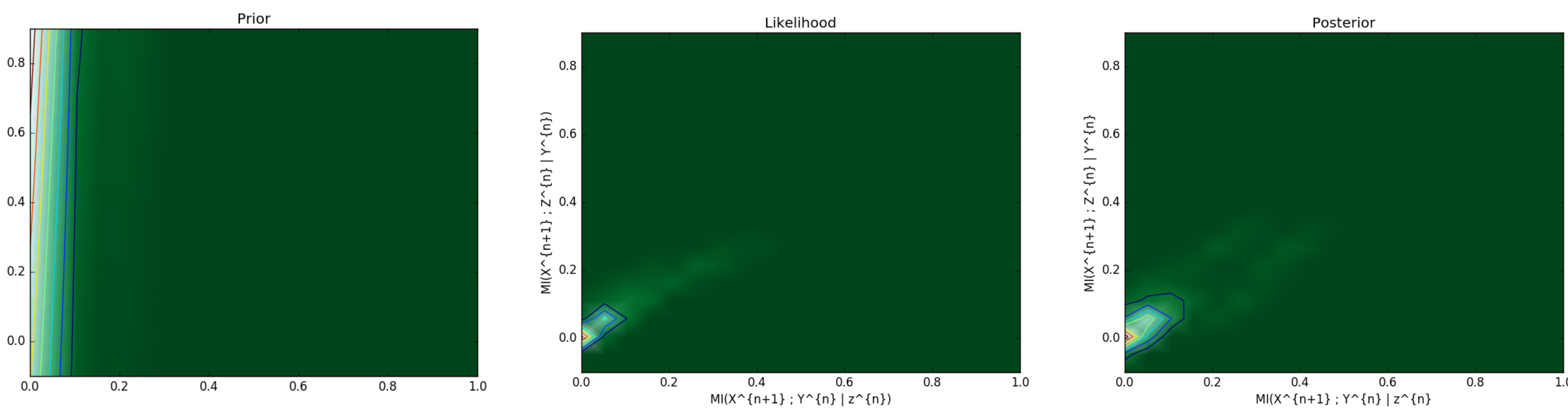
- If all causally relevant nodes or random processes are present, then **m** representing their collection or network is fixed and dependency drops out
- Back to usual Bayes' :
  $\Rightarrow p(CS \mid CS_{obs}) = p(CS)\, p(CS_{obs} \mid CS) \,/\, p(CS_{obs})$

Graphical Models : Structure Learning - David Heckerman, Microsoft Research
https://pdfs.semanticscholar.org/b883/049e29f168fab0d3279bebe2ed7381287472.pdf

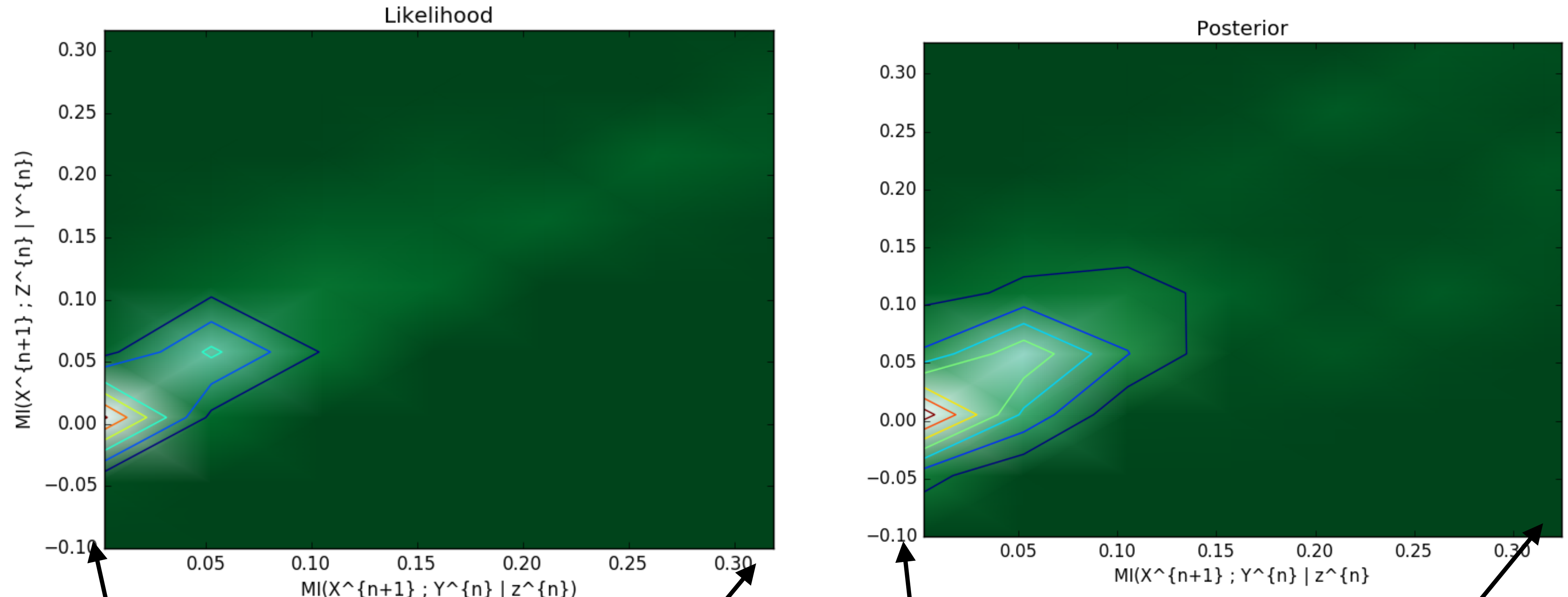
Toy Model I

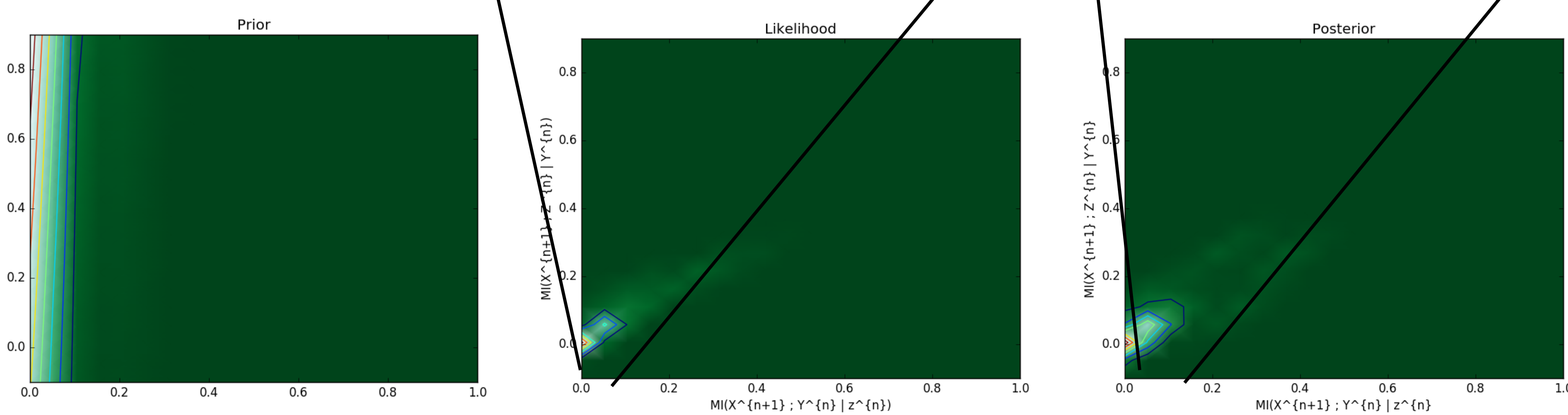
- $z^i = \eta z^i$
- $y^i = \eta\_ \; y^i$
- $x^{i+1} = 2\, y^i + z^i + \eta x^i$

Preliminary



- $z^i = \eta z^i$
- $y^i = \eta\_ \, y^i$
- $x^{i+1} = 2\, y^i \;+ z$





Conclusions

- Strength of causality relationships have uncertainties and therefore a distribution
- Attempt to estimate this from Bayesian framework
- When viewed as a graphical network with all relevant processes present, the estimation reduces to classic Bayes’ theorem
- Testing extensively with examples, this can be method can be used to estimate “causality state” upon new insights or observations without recomputing from scratch

Stay tuned…..