Introduction to continuous and hybrid Bayesian networks

Joanna Ficek

Supervisor: Paul Fink, M.Sc.

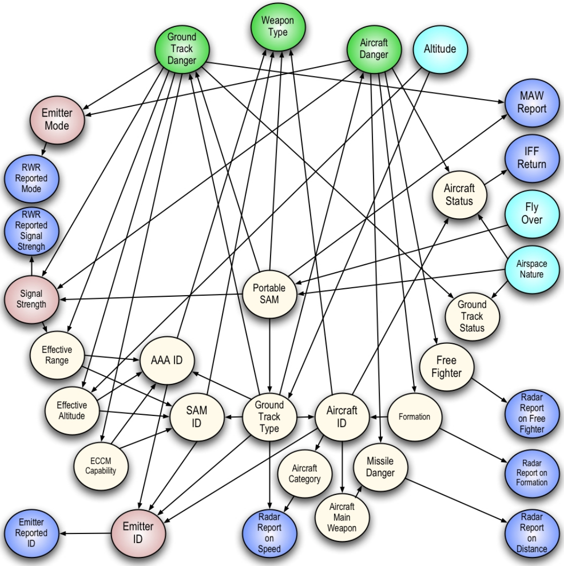
Department of Statistics

LMU

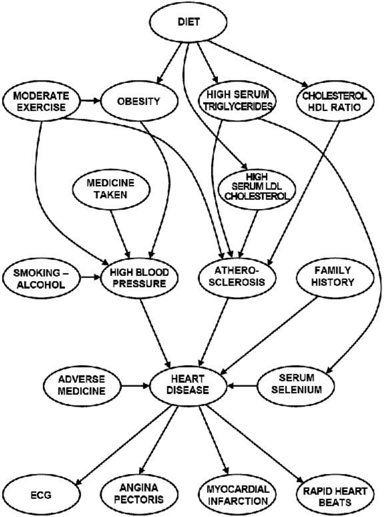
January 16, 2016

Outline

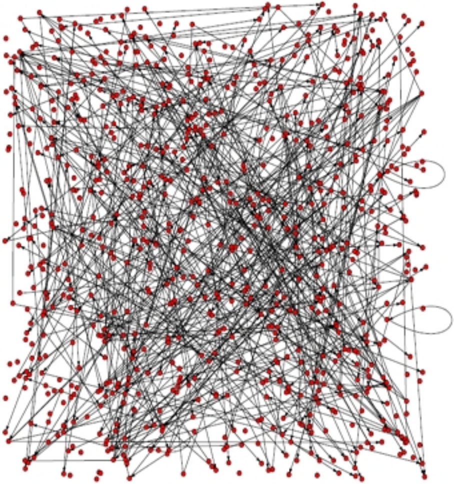
Why? Where?



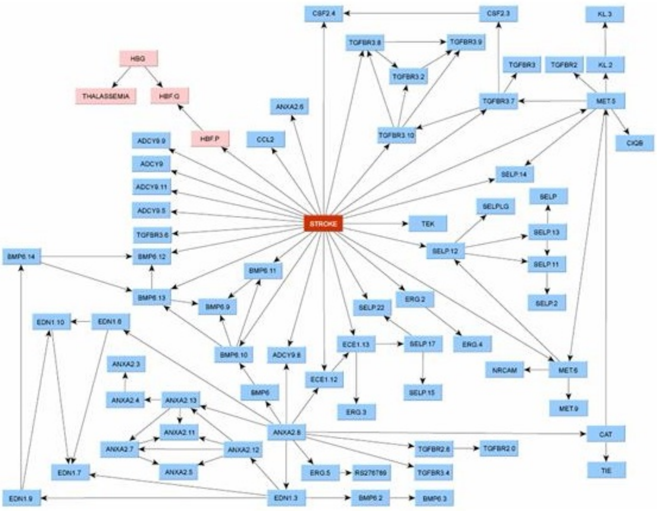
"Bayesian networks", http://www.pr-owl.org/ [access 15.01.2016]



Barbini et al. "Bayesian Approach in Medicine and Health Management"



Wenying Yan et al. "Effects of Time Point Measurement on the Reconstruction of Gene Regulatory Networks", http://www.mdpi.com/ [access 13.01.2016]



Sebastiani et al., Nature Genetics 37:435,2005

Applications

- medical diagnosis
- gene expression data
- complex genetic models
- robot localization
- risk management in robotics
- credit scoring
- ...

Continuous nodes

1. Challenges:

Continuous nodes

1. Challenges:
   - no representation for all possible densities
     (unlike CPTs in discrete BNs)
   - inference issues
   - complex distributions
2. Solutions:

Continuous nodes

1. Challenges:
   - no representation for all possible densities
     (unlike CPTs in discrete BNs)
   - inference issues
   - complex distributions
2. Solutions:
   - discretization
   - Linear Models
   - Gaussian approximation

Discretization

1. Idea: continuous domain into a finite set of intervals
2. Methods: Equal Interval Width, Equal Interval Frequency, . . .

Discretization

1. Idea: continuous domain into a finite set of intervals
2. Methods: Equal Interval Width, Equal Interval Frequency, ...

   Select $y^* \in [y^1, y^2]$

$$P(X \in [x^1, x^2] \mid y^*) = \int_{x^1}^{x^2} p(x \mid y^*) dy$$

Discretization

1. Idea: continuous domain into a finite set of intervals
2. Methods: Equal Interval Width, Equal Interval Frequency, . . .

Discretization

1. Idea: continuous domain into a finite set of intervals
2. Methods: Equal Interval Width, Equal Interval Frequency, ...
3. Limitations:
   - loss of information
   - trade-off: accuracy vs. computational cost $O(d^c)$

Discretization

1. Idea: continuous domain into a finite set of intervals
2. Methods: Equal Interval Width, Equal Interval Frequency, ...
3. Limitations:
   - loss of information
   - trade-off: accuracy vs. computational cost $O(d^c)$
4. Alternative: Linear Models

Linear Models

Broad class of models that satisfy independence of casual influence:

influence of multiple causes can be decomposed into separate influences.

Linear Models

Broad class of models that satisfy independence of casual influence:

influence of multiple causes can be decomposed into separate influences.

- the effect of parents $(Y_1, ..., Y_n)$ on X can be summarized via linear function

$$f(Y_1, ..., Y_n) = \sum_{i=0}^{k} w_i Y_i$$

  where w are coefficients.

- no interactions between Y_i 's (only through $f(Y_1, ..., Y_n)$)

Linear Gaussian model

Linear Gaussian model

Definition

Let X be a continuous variable with continuous parents $Y_1, ..., Y_k$. We say that X has a **Linear Gaussian** *CPD if there are $\beta_0, ..., \beta_k$ and σ^2 such that:*

$$p(X \mid \boldsymbol{y}) = \mathcal{N}(\beta_0 + \boldsymbol{\beta}^T \boldsymbol{y}, \sigma^2)$$

Linear Gaussian model

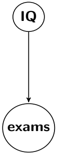
Definition

Let X be a continuous variable with continuous parents $Y_1, ..., Y_k$. We say that X has a **Linear Gaussian** *CPD if there are $\beta_0, ..., \beta_k$ and σ^2 such that:*

$$p(X \mid \boldsymbol{y}) = \mathcal{N}(\beta_0 + \boldsymbol{\beta}^T \boldsymbol{y}, \sigma^2)$$

$X = \beta_0 + \beta_1 y_1 + ... + \beta_k y_k + \epsilon$

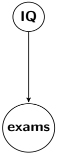
Linear Gaussian model



$$p(IQ) = \mathcal{N}(100, 15)$$

$$p(E \mid IQ) = \mathcal{N}(15 + 0.6IQ, 10) = \mathcal{N}(75, 10)$$

Linear Gaussian model



$$p(IQ) = \mathcal{N}(100, 15)$$

$$p(E \mid IQ) = \mathcal{N}(15 + 0.6IQ, 10) = \mathcal{N}(75, 10)$$

Definition

A **Gaussian Bayesian network** *is a Bayesian network where all the variables are continuous and where CPDs are linear Gaussians.*

Gaussian BN $\Rightarrow$ joint Gaussian

Theorem

Let X be a linear Gaussian of its parents $Y_1, ..., Y_k$: $p(X \mid \boldsymbol{y}) = \mathcal{N}(\beta_0 + \boldsymbol{\beta}^T \boldsymbol{y}; \sigma^2)$

Assume, that $Y_1, ..., Y_k$ are jointly Gaussian with distribution $\mathcal{N}(\boldsymbol{\mu}; \boldsymbol{\Sigma})$. Then:

- *The distribution of X is a normal distribution $p(X) = \mathcal{N}(\mu_X; \sigma_X^2)$ where:*

$$\mu_X = \beta_0 + \boldsymbol{\beta}^T \boldsymbol{\mu} \qquad \sigma_X^2 = \sigma^2 + \boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta}$$

- *The joint distribution over $\{\boldsymbol{Y}, X\}$ is a normal distribution, where:*

$$Cov[Y_i, X] = \sum_{j=0}^{k} \beta_j \boldsymbol{\Sigma_{i,j}}$$

Gaussian BN $\Rightarrow$ joint Gaussian

Theorem

Let X be a linear Gaussian of its parents $Y_1, ..., Y_k$: $p(X \mid \boldsymbol{y}) = \mathcal{N}(\beta_0 + \boldsymbol{\beta}^T \boldsymbol{y}; \sigma^2)$

Assume, that $Y_1, ..., Y_k$ are jointly Gaussian with distribution $\mathcal{N}(\boldsymbol{\mu}; \boldsymbol{\Sigma})$. Then:

- *The distribution of X is a normal distribution $p(X) = \mathcal{N}(\mu_X; \sigma_X^2)$ where:*

$$\mu_X = \beta_0 + \boldsymbol{\beta}^T \boldsymbol{\mu} \qquad \sigma_X^2 = \sigma^2 + \boldsymbol{\beta}^T \boldsymbol{\Sigma} \boldsymbol{\beta}$$

- *The joint distribution over $\{\boldsymbol{Y}, X\}$ is a normal distribution, where:*

$$Cov[Y_i, X] = \sum_{j=0}^{k} \beta_j \boldsymbol{\Sigma_{i,j}}$$

$\Rightarrow$ A Gaussian Bayesian network defines a joint Gaussian distribution.

Gaussian BN $\Leftarrow$ joint Gaussian

Theorem

Let $\mathcal{X} = X_1, \ldots, X_n$ and let P be a joint Gaussian distribution over $\mathcal{X}$.

Given any ordering $X_1, \ldots, X_n$ over $\mathcal{X}$, we can construct a Bayesian network graph $\mathcal{G}$ and a Bayesian network $\mathcal{B}$ such that:

1. $Pa_i^{\mathcal{G}} \subseteq X_1, \ldots, X_{i-1}$;
2. *the CPD of X_i in $\mathcal{B}$ is a linear Gaussian of its parents;*
3. *$\mathcal{G}$ is a minimal I-map for P.*

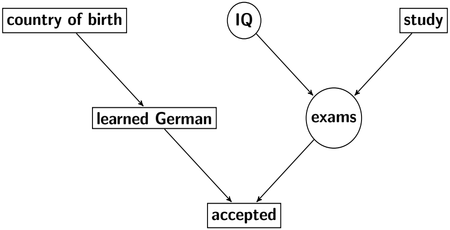
Probability of being accepted to a german university

(MA programme) for international students form other EU-countries.

accepted

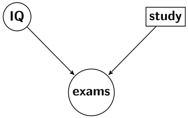
Probability of being accepted to a german university

(MA programme) for international students form other EU-countries.



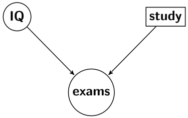
Hybrid BNs

- continuous children with discrete (and continuous) parents

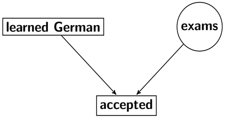


Hybrid BNs

- continuous children with discrete (and continuous) parents



- discrete children with continuous (and discrete) parents



Continuous children with discrete parents

Continuous children with discrete parents

Definition

Let X be a continuous variable with discrete $\mathbf{A} = A_1, ..., A_m$ and continuous $\mathbf{Y} = Y_1, ..., Y_k$ parents. We define the **Conditional Linear Gaussian** *(CLG) model as:*

$$p(X \mid \boldsymbol{a}, \boldsymbol{y}) = \mathcal{N}(w_{\boldsymbol{a},0} + \sum_{i=1}^{k} w_{\boldsymbol{a},i} y_i; \sigma_a^2)$$

where w are coefficients.

Continuous children with discrete parents

Definition

Let X be a continuous variable with discrete $\mathbf{A} = A_1, ..., A_m$ *and continuous* $\mathbf{Y} = Y_1, ..., Y_k$ *parents. We define the* **Conditional Linear Gaussian** *(CLG) model as:*

$$p(X \mid \boldsymbol{a}, \boldsymbol{y}) = \mathcal{N}(w_{\boldsymbol{a},0} + \sum_{i=1}^{k} w_{\boldsymbol{a},i} y_i; \sigma_a^2)$$

where w are coefficients.

- separate linear Gaussian model for each assignment to discrete parents

Continuous children with discrete parents

Definition

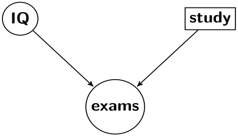
Let X be a continuous variable with discrete $\mathbf{A} = A_1, ..., A_m$ and continuous $\mathbf{Y} = Y_1, ..., Y_k$ parents. We define the **Conditional Linear Gaussian** *(CLG) model as:*

$$p(X \mid \boldsymbol{a}, \boldsymbol{y}) = \mathcal{N}(w_{\boldsymbol{a},0} + \sum_{i=1}^{k} w_{\boldsymbol{a},i} y_i ; \sigma_a^2)$$

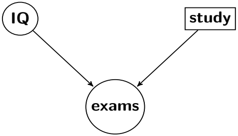
where w are coefficients.

- separate linear Gaussian model for each assignment to discrete parents
- defines a conditional Gaussian joint distribution [Lerner et al.,2001]

Continuous children with discrete parents



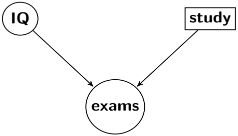
Continuous children with discrete parents



$p(IQ) = \mathcal{N}(100, 15)$

$p(E \mid IQ, S = s^1) = \mathcal{N}(25 + 0.6IQ, 10) = \mathcal{N}(85, 10)$

Continuous children with discrete parents



$p(IQ) = \mathcal{N}(100, 15)$

$p(E \mid IQ, S = s^1) = \mathcal{N}(25 + 0.6IQ, 10) = \mathcal{N}(85, 10)$

$p(E \mid IQ, S = s^0) = \mathcal{N}(-5 + 0.7IQ, 12) = \mathcal{N}(65, 12)$

Discrete children with continuous parents

Discrete children with continuous parents

- Threshold (τ) which determines the change in discrete values

Discrete children with continuous parents

- Threshold (τ) which determines the change in discrete values

X binary $\{x^0, x^1\}$ with parents $Y_1, ..., Y_k$:

$f(Y_1, ..., Y_k) \geq \tau \Rightarrow P(X = x^1)$ likely to be 1

$f(Y_1, ..., Y_k) < \tau \Rightarrow P(X = x^1)$ likely to be 0

Discrete children with continuous parents

- Threshold (τ) which determines the change in discrete values

X binary $\{x^0, x^1\}$ with parents $Y_1, ..., Y_k$:

$f(Y_1, ..., Y_k) \geq \tau \Rightarrow P(X = x^1)$ likely to be 1

$f(Y_1, ..., Y_k) < \tau \Rightarrow P(X = x^1)$ likely to be 0

$f(Y_1, ..., Y_k) = w_0 + \sum_{i=1}^{k} w_i Y_i$

Discrete children with continuous parents

- Threshold (τ) which determines the change in discrete values

X binary $\{x^0, x^1\}$ with parents $Y_1, ..., Y_k$:

$f(Y_1, ..., Y_k) \geq \tau \Rightarrow P(X = x^1)$ likely to be 1

$f(Y_1, ..., Y_k) < \tau \Rightarrow P(X = x^1)$ likely to be 0

$f(Y_1, ..., Y_k) = w_0 + \sum_{i=1}^{k} w_i Y_i$

Definition

The CPD $P(X|Y_1, ..., Y_k)$ is a logistic CPD if there are $k+1$ weights $w_0, w_1, ..., w_k$ such that:

$$P(x^1|Y_1, ..., Y_k) = sigmoid(w_0 + \sum_{i=1}^{k} w_i Y_i)$$

Discrete children with continuous parents

- Threshold which determines the change in discrete values
- Augmented CLGs [Lerner et al. (2001)]

Definition

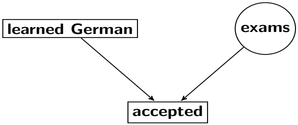
Let A be a discrete variable with possible values $a_1, ..., a_m$ and let $\mathbf{Y} = Y_1, ..., Y_k$ denote its continuous parents. We define the CPD in **augmented Conditional Linear Gaussian** *model as:*

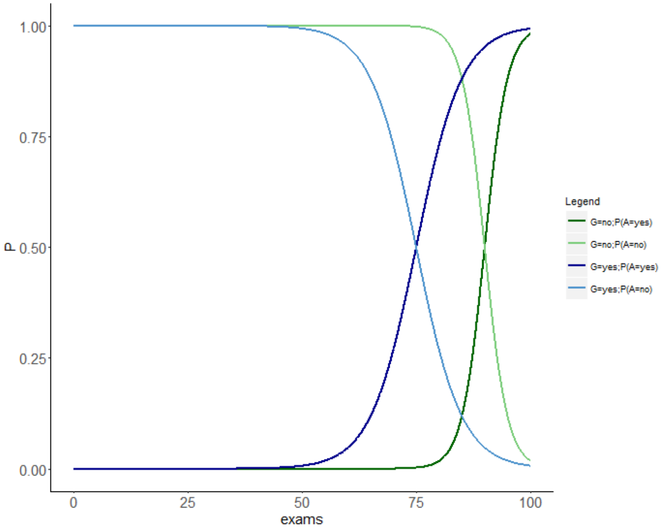
$$p(A = a_i \mid y_1, ..., y_k) = \frac{exp(w_{i,0} + \sum_{l=1}^{k} w_{i,l} y_l)}{\sum_{j=1}^{m} exp(w_{j,0} + \sum_{s=1}^{k} w_{j,s} y_s)}$$

Discrete children with continuous parents

Definition

A Bayesian network with all discrete variables having only discrete parents and continuous variables having a CLG CPD is a **Conditional Linear Gaussian network***.*




1.00
0.75
0.50
0.25
0.00
P
0
25
50
75
100
exams
Legend
G=no;P(A=yes)
G=no;P(A=no)
G=yes;P(A=yes)
G=yes;P(A=no)

Exponential family

An exponential family is specified by:

- a sufficient statistics function τ
- a parameter space that is a convex set Θ of legal parameters
- a natural parameter function t
- an auxiliary measure A over $\mathcal{X}$

Exponential family

An exponential family is specified by:

- a sufficient statistics function τ
- a parameter space that is a convex set Θ of legal parameters
- a natural parameter function t
- an auxiliary measure A over $\mathcal{X}$

Definition

An exponential family $\mathcal{P} = \{P_\theta : \boldsymbol{\theta} \in \Theta\}$ over set of variables $\mathcal{X}$, where

$$P_\theta(\xi) = \frac{1}{Z(\theta)} A(\xi) exp\{\langle t(\boldsymbol{\theta}), \tau(\xi)\rangle\}$$

with partition function

$$Z(\boldsymbol{\theta}) = \sum_{\xi} A(\xi) exp\{\langle t(\boldsymbol{\theta}), \tau(\xi)\rangle\}$$

Univariate normal distribution

$$\tau(x) = \langle x, x^2 \rangle \quad (1)$$

$$t(\mu, \sigma^2) = \langle \frac{\mu}{\sigma^2}, -\frac{1}{2\sigma^2} \rangle \quad (2)$$

$$Z(\mu, \sigma^2) = \sqrt{2\pi}\sigma exp\left\{ \frac{\mu^2}{2\sigma^2} \right\} \quad (3)$$

Then,

$$P(x) = \frac{1}{Z(\mu, \sigma^2)} exp\{\langle t(\theta), \tau(X)\} = \frac{1}{\sqrt{2\pi}\sigma} exp\left\{ -\frac{(x-\mu)^2}{2\sigma^2} \right\}$$

Product distribution

Definition

A factor in an exponential family: $\phi_{\boldsymbol{\theta}}(\xi) = A(\xi) exp\{\langle t(\boldsymbol{\theta}), \tau(\xi)\rangle\}$

Definition

Let $\Phi_1, ..., \Phi_k$ *be exponential factor families. The composition of* $\Phi_1, ..., \Phi_k$ *is the family* $\Phi_1 \times \Phi_2 \times ... \times \Phi_k$ *parametrized by* $\boldsymbol{\theta} = \boldsymbol{\theta}_1 \circ \boldsymbol{\theta}_2 \circ ... \circ \boldsymbol{\theta}_k \in \Theta_1 \times \Theta_2 \times ... \times \Theta_k$ *:*

$$P_\theta \propto \prod_i \phi_{\boldsymbol{\theta}_i}(\xi) = \left(\prod_i A_i(\xi)\right) exp\left\{\sum_i \langle t(\boldsymbol{\theta}), \tau(\xi)\rangle\right\}$$

Product distribution

Definition

A factor in an exponential family: $\phi_{\boldsymbol{\theta}}(\xi) = A(\xi)exp\{\langle t(\boldsymbol{\theta}), \tau(\xi)\rangle\}$

Definition

Let $\Phi_1, ..., \Phi_k$ *be exponential factor families. The composition of* $\Phi_1, ..., \Phi_k$ *is the family* $\Phi_1 \times \Phi_2 \times ... \times \Phi_k$ *parametrized by* $\boldsymbol{\theta} = \boldsymbol{\theta}_1 \circ \boldsymbol{\theta}_2 \circ ... \circ \boldsymbol{\theta}_k \in \Theta_1 \times \Theta_2 \times ... \times \Theta_k$ *:*

$$P_\theta \propto \prod_i \phi_{\boldsymbol{\theta}_i}(\xi) = \left(\prod_i A_i(\xi)\right) exp\left\{\sum_i \langle t(\boldsymbol{\theta}), \tau(\xi)\rangle\right\}$$

A Bayesian network with locally normalized exponential CPDs defines an exponential family.

Product distribution

Definition

A factor in an exponential family: $\phi_{\boldsymbol{\theta}}(\xi) = A(\xi) exp\{\langle \boldsymbol{\theta}, \tau(\xi) \rangle\}$

Definition

Let $\Phi_1, ..., \Phi_k$ *be exponential factor families. The composition of* $\Phi_1, ..., \Phi_k$ *is the family* $\Phi_1 \times \Phi_2 \times ... \times \Phi_k$ *parametrized by* $\boldsymbol{\theta} = \boldsymbol{\theta}_1 \circ \boldsymbol{\theta}_2 \circ ... \circ \boldsymbol{\theta}_k \in \Theta_1 \times \Theta_2 \times ... \times \Theta_k$ *:*

$$P_\theta \propto \prod_i \phi_{\boldsymbol{\theta}_i}(\xi) = \left(\prod_i A_i(\xi)\right) exp\left\{\sum_i \langle t_i(\boldsymbol{\theta}_i), \tau_i(\xi) \rangle\right\}$$

A Bayesian network with locally normalized exponential CPDs defines an exponential family.

Entropy

Measure of degree of disorder in a system (thermodynamics, R. Clausius, 1865)

Shannon's Measure of Uncertainty - statistical entropy:

Definition

Let $P(X)$ be a distribution over a random variable X. The entropy of X is

$$\boldsymbol{H}_P(X) = -\boldsymbol{E}_P[log P(X)]$$

Entropy

Measure of degree of disorder in a system (thermodynamics, R. Clausius, 1865)

Shannon's Measure of Uncertainty - statistical entropy:

Definition

Let $P(X)$ be a distribution over a random variable X. The entropy of X is

$$\boldsymbol{H}_P(X) = -\boldsymbol{E}_P[log P(X)]$$

small entropy ⇒ probability mass on a few instances

large entropy ⇒ probability mass uniformly spread

Entropy

Theorem

Let P_θ be a distribution in an exponential family defined by the functions τ and t.

Then

$$\boldsymbol{H}_{P_\theta}(X) = \ln Z(\boldsymbol{\theta}) - \langle \boldsymbol{E}_{P_\theta}[\tau(\mathcal{X})], t(\boldsymbol{\theta})\rangle$$

Entropy

Theorem

Let P_θ be a distribution in an exponential family defined by the functions τ and t.

Then

$$\boldsymbol{H}_{P_\theta}(X) = lnZ(\boldsymbol{\theta}) - \langle \boldsymbol{E}_{P_\theta}[\tau(\mathcal{X})], t(\boldsymbol{\theta}) \rangle$$

Theorem

Let $P(\mathcal{X}) = \prod_i P(X_i \mid Pa_i^{\mathcal{G}})$ be a distribution consistent with a Bayesian network $\mathcal{G}$.

Then

$$\boldsymbol{H}_P(\mathcal{X}) = \sum_i \boldsymbol{H}_P(X_i \mid Pa_i^{\mathcal{G}}) = \sum_i \sum_{pa_i^{\mathcal{G}}} P(pa_i^{\mathcal{G}}) \boldsymbol{H}_P(X_i \mid pa_i^{\mathcal{G}})$$

Relative entropy

Complex distribution ⇒ approximation

Measure of inaccuracy: relative entropy (Kullback-Leibler distance)

Definition

Let Q and P be two distributions over random variables $X_1, ..., X_n$.

The relative entropy of Q and P is:

$$\boldsymbol{D}(Q(X_1, ..., X_n) \parallel P(X_1, ..., X_n)) = \boldsymbol{E}_Q\left[log \frac{Q(X_1, ..., X_n)}{P(X_1, ..., X_n)}\right]$$

where we set log(0) = 0.

Relative entropy

Complex distribution $\Rightarrow$ approximation

Measure of inaccuracy: relative entropy (Kullback-Leibler distance)

Definition

Let Q and P be two distributions over random variables $X_1, ..., X_n$.

The relative entropy of Q and P is:

$$\boldsymbol{D}(Q(X_1, ..., X_n) \parallel P(X_1, ..., X_n)) = \boldsymbol{E}_Q \left[log \frac{Q(X_1, ..., X_n)}{P(X_1, ..., X_n)} \right]$$

where we set log(0) = 0.

- $\bigvee_{P,Q} \boldsymbol{D}(Q \parallel P) \geq 0$
- $\boldsymbol{D}(Q \parallel P)$ small $\Rightarrow$ P close to Q $\Rightarrow$ small loss of information
- $\bigvee_{P \neq Q} \boldsymbol{D}(Q \parallel P) \neq D(P \parallel Q)$

Relative entropy

Theorem

Consider a distribution Q and a distribution P_θ in an exponential family defined by τ and t. Then

$$\boldsymbol{D}(Q \parallel P_\theta) = -\boldsymbol{H}_Q(\mathcal{X}) - \langle \boldsymbol{E}_Q[\tau(\mathcal{X})], t(\boldsymbol{\theta}) \rangle + lnZ(\boldsymbol{\theta})$$

Relative entropy

Theorem

Consider a distribution Q and a distribution P_θ in an exponential family defined by τ and t. Then

$$\boldsymbol{D}(Q \parallel P_\theta) = -\boldsymbol{H}_Q(\mathcal{X}) - \langle \boldsymbol{E}_Q[\tau(\mathcal{X})], t(\boldsymbol{\theta}) \rangle + lnZ(\boldsymbol{\theta})$$

Theorem

If P and Q are distributions over $\mathcal{X}$ consistent with a Bayesian network $\mathcal{G}$, then

$$\boldsymbol{D}(Q \parallel P) = \sum_i \sum_{pa_i^{\mathcal{G}}} Q(pa_i^{\mathcal{G}}) \boldsymbol{D}(Q(X_i, pa_i^{\mathcal{G}})) \parallel P(X_i \mid pa_i^{\mathcal{G}})$$

Projections

Project a distribution P onto family of distributions $\mathcal{Q}$.

- I-projections

$$Q^I = arg \min_{Q \in \mathcal{Q}} \boldsymbol{D}(Q \parallel P)$$

- M-projections

$$Q^M = arg \min_{Q \in \mathcal{Q}} \boldsymbol{D}(P \parallel Q)$$

In general: $Q^I \neq Q^M$

I-projections

$$Q^I = arg \min_{Q \in \mathcal{Q}} \boldsymbol{D}(Q \parallel P) = arg \min_{Q \in \mathcal{Q}} (-\boldsymbol{H}_Q(X) + \boldsymbol{E}_Q[-lnP(X)])$$

- if complex graphical model
- high density where P is large

  low density where P is small
- penalty for low entropy
- some simplification of computation is possible

M-projections

$$Q^M = arg \min_{Q \in \mathcal{Q}} \boldsymbol{D}(P \parallel Q) = arg \min_{Q \in \mathcal{Q}} (-\boldsymbol{H}_P(X) + \boldsymbol{E}_P[-lnQ(X)])$$

- learning problem
- attempts to match the main mass of P:
  - high density to the regions probable according to P
  - high penalty for low density in these regions
- relatively high variance
- use of exponential form of the distribution may simplify the computation

M-projections

Let P be a distribution over $\mathcal{X}$.

Theorem

Let $\mathcal{Q}$ be an exponential family defined by τ and t. Then $Q^M = Q_\theta$ if there is a set of parameters $\boldsymbol{\theta}$ such that

$$E_{Q_\theta}[\tau(\mathcal{X})] = E_P[\tau(\mathcal{X})]$$

M-projections

Let P be a distribution over $\mathcal{X}$.

Theorem

Let $\mathcal{Q}$ be an exponential family defined by τ and t. Then $Q^M = Q_\theta$ if there is a set of parameters $\boldsymbol{\theta}$ such that

$$E_{Q_\theta}[\tau(\mathcal{X})] = E_P[\tau(\mathcal{X})]$$

Theorem

Let $\mathcal{G}$ be a Bayesian network structure. Then

$$Q^M(\mathcal{X}) = \prod_i P(X_i \mid Pa_i^{\mathcal{G}})$$

Summary

1. Continuous and hybrid BNs have many applications
2. Challenges: representation, inference
3. Solutions: discretization, Linear Models, approximation
4. Exponential family - useful form
5. True distribution unknown or complex

   $\Rightarrow$ entropy, relative entropy
6. Projections - find approximation

Bibliography

1. Koller D. & Friedman N., 2009, *Probabilistic Graphical Models. Principles and Techniques.*, The MIT Press, Massachusetts, USA
2. Koller D., 2013, on-line course "Probabilistic Graphical Models", https://class.coursera.org/pgm/lecture
3. Lerner et al., 2001, *Exact Inference in Networks with Discrete Children of Continuous Parent*, p.319-328, UAI 2001
4. Pourret et al.,2008, *Bayesian networks : a practical guide to applications*, John Wiley & Sons Ltd, ISBN: 978-0-470-06030-8, online: http://bayanbox.ir/view/1741861298367825388/Olivier-Pourret-Patrick-Na-Bruce-Marcot-Bay.pdf

Bibliography

1. Friedman et al., 2000, *Using Bayesian Networks to Analyze Expression Data*, Journal of Computational Biology, Volume7, No. 3/4
2. Lauritzen S. & Sheehan N., 2003, *Graphical Model for Genetic Analyses*, Statistical Science 2003, Vol. 18, No. 4, 489514

Thank you for your attention!