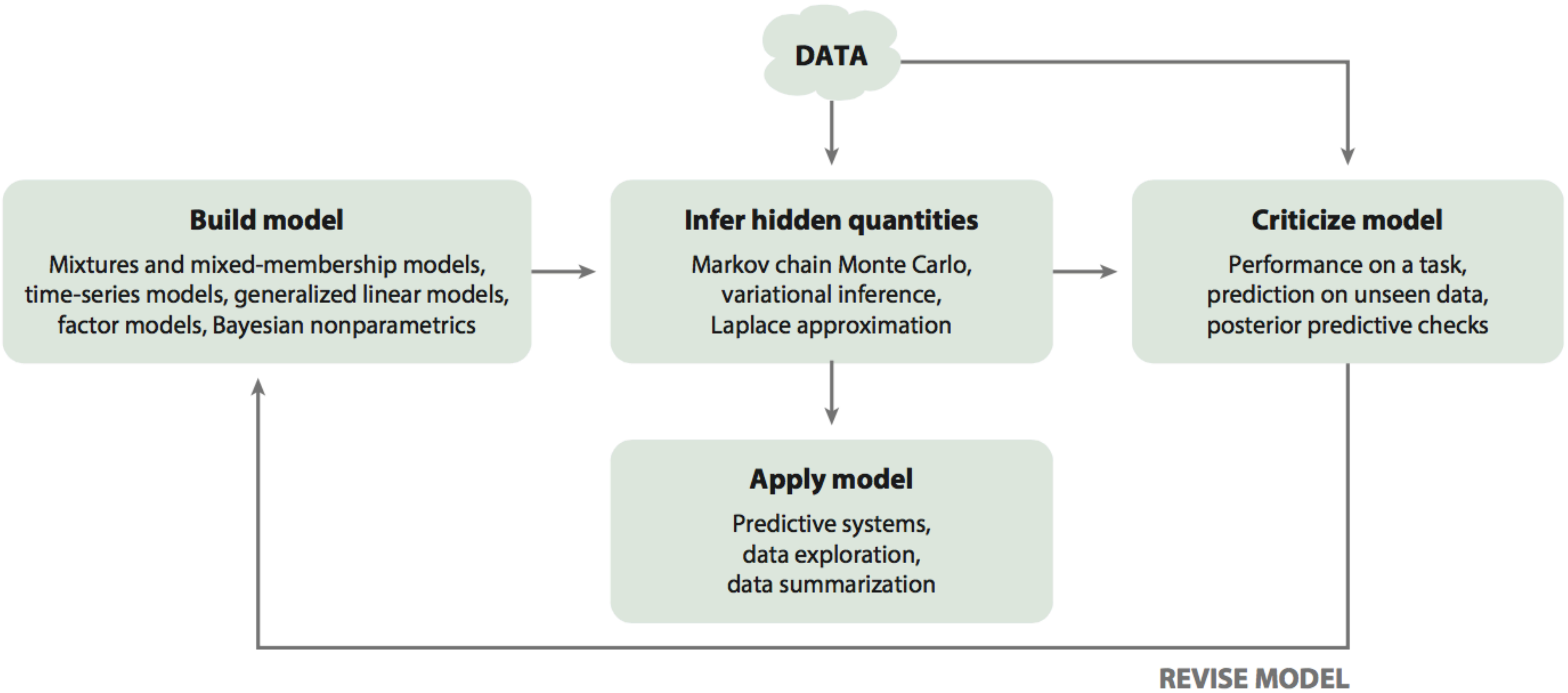
AMiDST TOOLBOX

Latent Variable Models

Andrés R. Masegosa
University of Almeria
andres.masegosa@ual.es

January, 2018 Geilo (Norway)

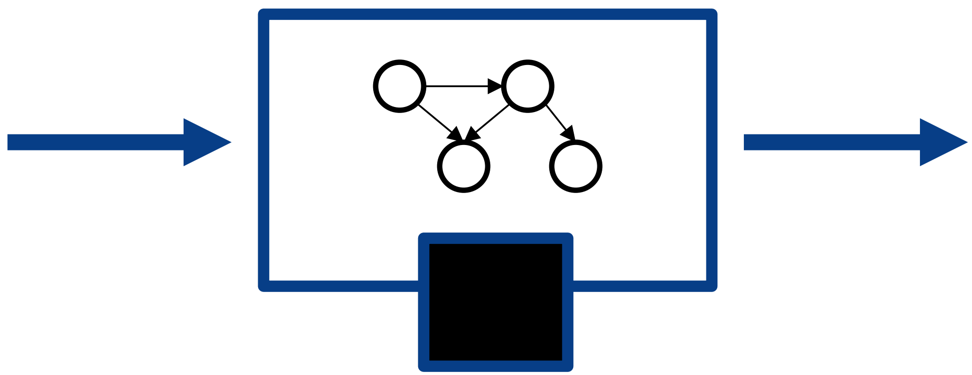
BOX'S LOOP



Blei, David M. "Build, compute, critique, repeat: Data analysis with latent variable models." *Annual Review of Statistics and Its Application* 1 (2014): 203-232.

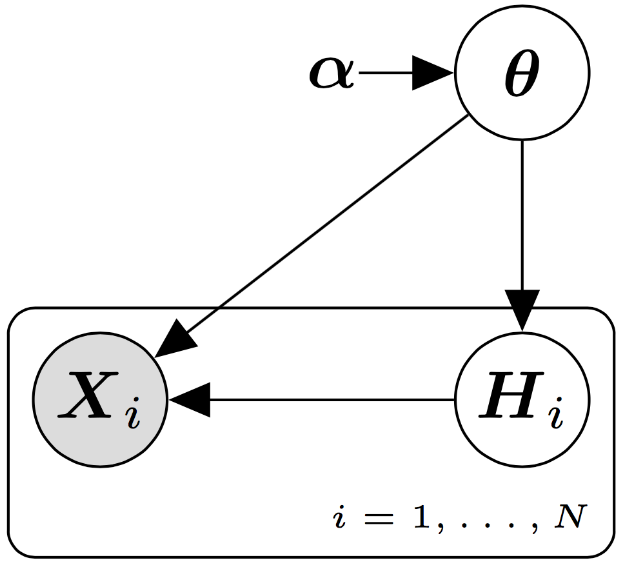
Openbox Models
[Probabilistic Graphical Models]

[Big] Data
[+Prior Information]



Knowledge
[+ Predictions]

[Scalable] Bayesian Inference Engine
(Powered by Variational Methods)



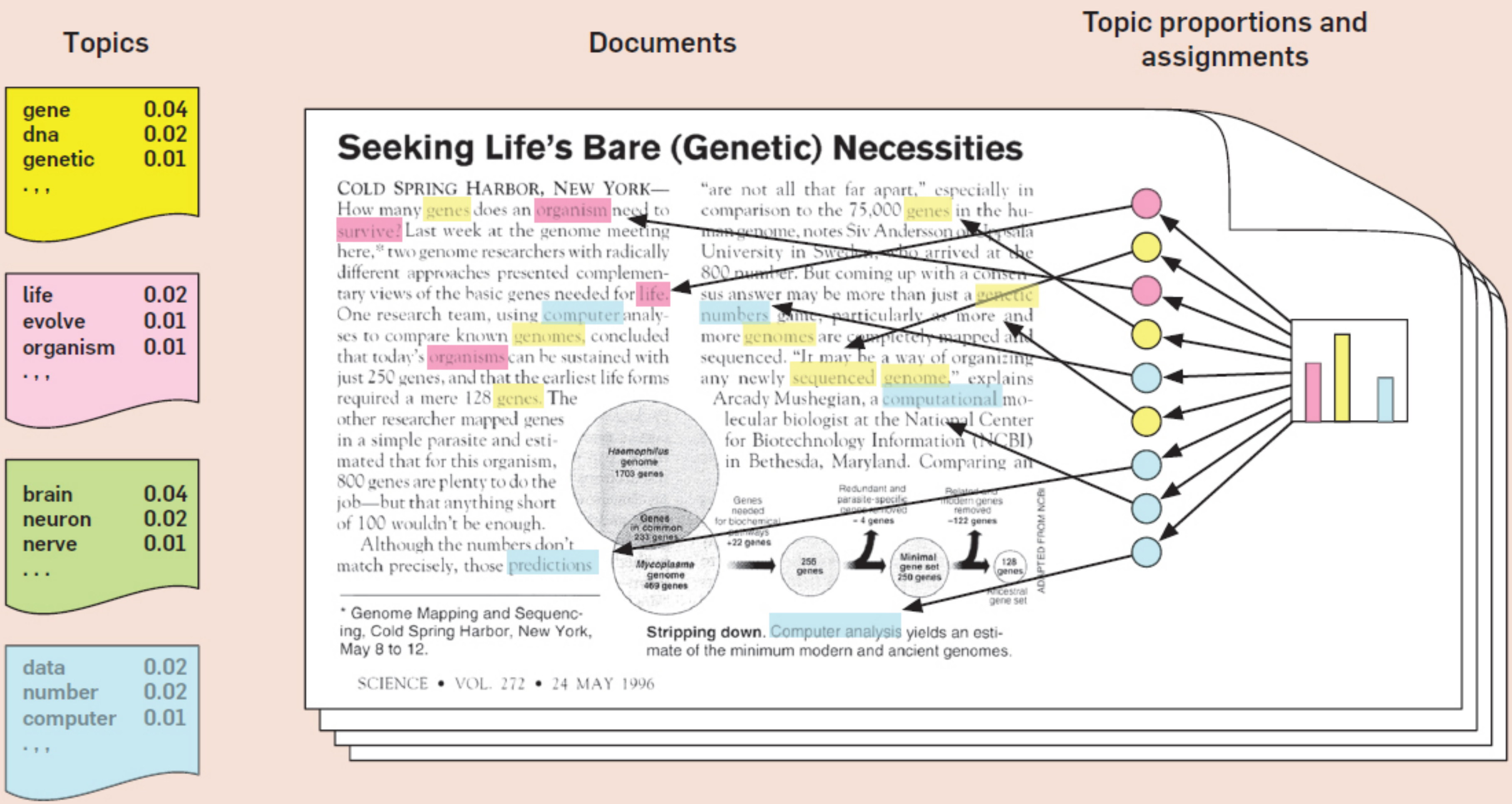
$$p(\theta, \mathbf{H} | D)$$

Latent Variable Models

Modeling non-observable mechanisms.

Text Modeling

AMiDST TOOLBOX



David Blei, Probabilistic Topic Models, Communications of the ACM, Vol. 55 No. 4, Pages 77-84

1	2	3	4	5
Game	Life	Film	Book	Wine
Season	Know	Movie	Life	Street
Team	School	Show	Books	Hotel
Coach	Street	Life	Novel	House
Play	Man	Television	Story	Room
Points	Family	Films	Man	Night
Games	Says	Director	Author	Place
Giants	House	Man	House	Restaurant
Second	Children	Story	War	Park
Players	Night	Says	Children	Garden

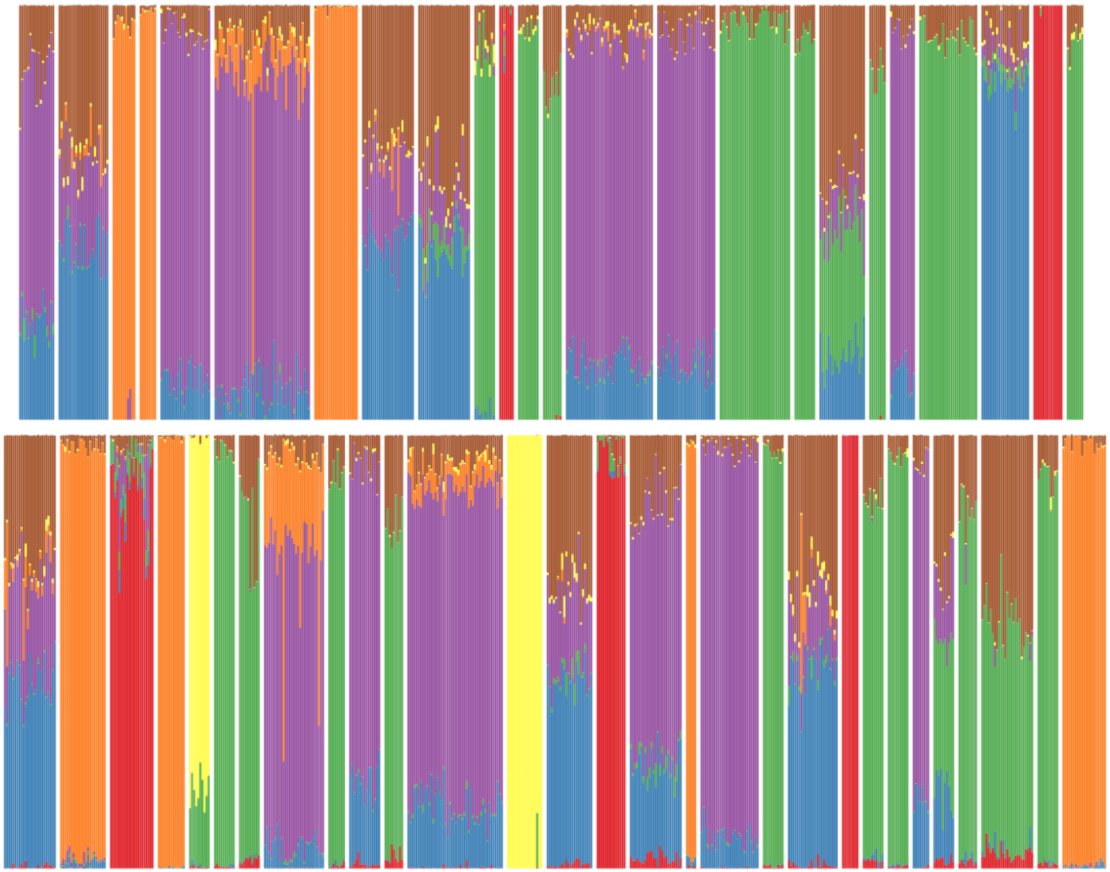
6	7	8	9	10
Bush	Building	Won	Yankees	Government
Campaign	Street	Team	Game	War
Clinton	Square	Second	Mets	Military
Republican	Housing	Race	Season	Officials
House	House	Round	Run	Iraq
Party	Buildings	Cup	League	Forces
Democratic	Development	Open	Baseball	Iraqi
Political	Space	Game	Team	Army
Democrats	Percent	Play	Games	Troops
Senator	Real	Win	Hit	Soldiers

11	12	13	14	15
Children	Stock	Church	Art	Police
School	Percent	War	Museum	Yesterday
Women	Companies	Women	Show	Man
Family	Fund	Life	Gallery	Officer
Parents	Market	Black	Works	Officers
Child	Bank	Political	Artists	Case
Life	Investors	Catholic	Street	Found
Says	Funds	Government	Artist	Charged
Help	Financial	Jewish	Paintings	Street
Mother	Business	Pope	Exhibition	Shot

Topics found in 1.8M articles from the New York Times

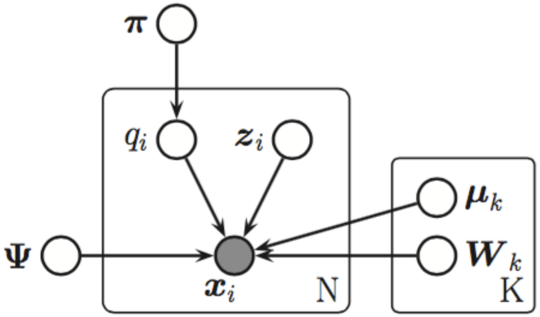
[Hoffman, Blei, Wang, Paisley, JMLR 2013]

Population Genetics



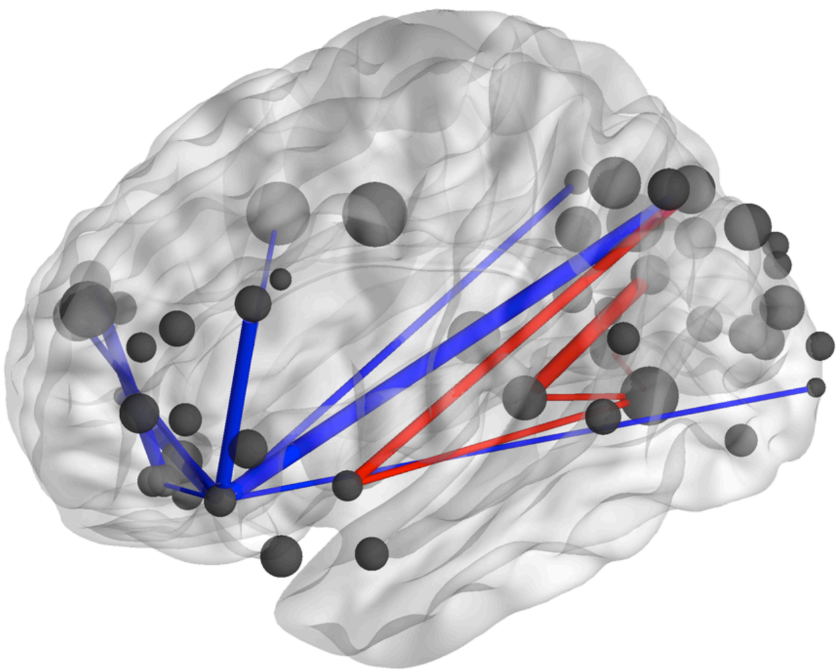
$$\beta_{k,\ell} \sim \text{Beta}(a, b)$$
$$\theta_i \sim \text{Dirichlet}(c)$$
$$x_{i,l} \sim \text{Binomial}\left(2, \textstyle\sum_k \theta_{i,k}\beta_{k,\ell}\right)$$

Gopalan, Prem, et al. Scaling probabilistic models of genetic variation to millions of humans. Nature Research, 2016.



Trun et al. Automatic Differentiation Variational Inference. JMLR, 2016.

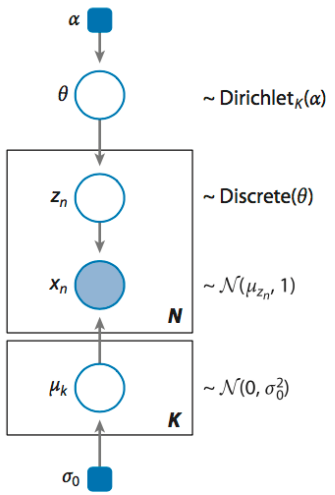
fMRI Data



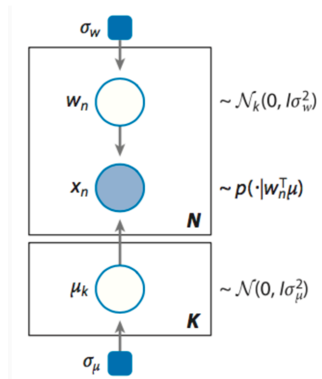
Neuroscience analysis of 220 million fMRI measurements

[Manning et al., PLOS ONE 2014]

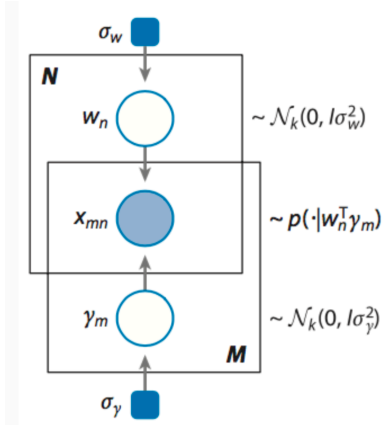
Blei, David M. "Build, compute, critique, repeat: Data analysis with latent variable models." *Annual Review of Statistics and Its Application* 1 (2014): 203-232.



Gaussian Mixture



Principal Component Analysis



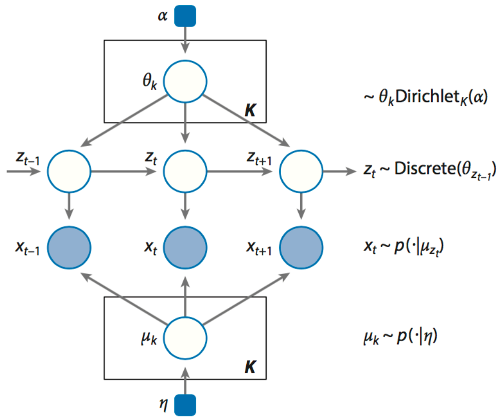
Matrix Factorization

Latent Variable Models

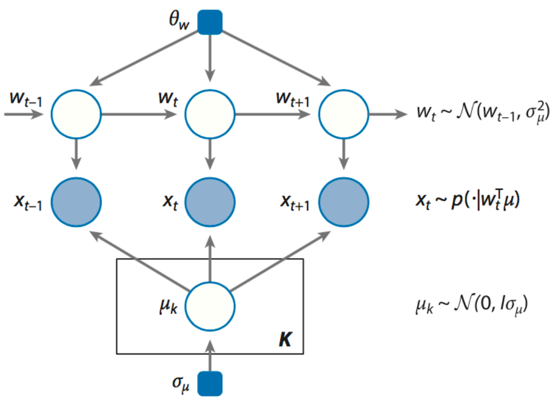
Gaussian Mixture Models, Principal Component Analysis, Factor Analyzers, Latent Dirichlet Allocation, etc.

AMIDST TOOLBOX

Blei, David M. "Build, compute, critique, repeat: Data analysis with latent variable models." *Annual Review of Statistics and Its Application* 1 (2014): 203-232.



Hidden Markov Model



Kalman Filter

Dynamic/Temporal Models

Hidden Markov Models, Linear Dynamical Systems, State Space Models, Input-Output HMM, etc.

LVMs in AMIDST

Automatic Bayesian Treatment

Modeling non-observable mechanisms.

$$P(\boxed{\theta \mid \mathbf{D}})$$

Bayesian Learning

STANDARD APPROACH

$$\{\theta : \theta \in \Theta\}$$

[Set of Models]

Data $\longrightarrow$ $\boxed{\arg\min_{\theta \in \Theta} \ell(D, \theta)}$ $\longrightarrow$ $\theta^\star$

[Optimal Model]

Loss Minimization

(Stochastic Gradient Descent)

Example: $y = \theta_0 + \theta_1 \cdot x_1 + \ldots + \theta_k \cdot x_k$

$$p(\theta)$$

[Prior over Models]

Data

$$\boxed{\frac{p(D_1|\theta)p(\theta)}{p(D_1)}}$$

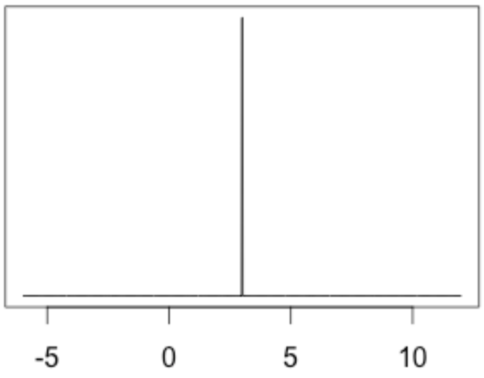
$$p(\theta|D_1)$$

[Posterior over Models]

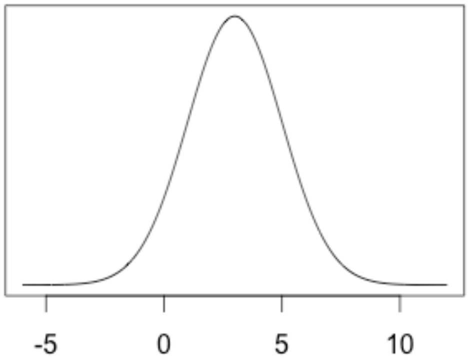
Bayes Theorem

Learning as an inference Problem

MODEL UNCERTAINTY



vs



$$\theta^\star$$

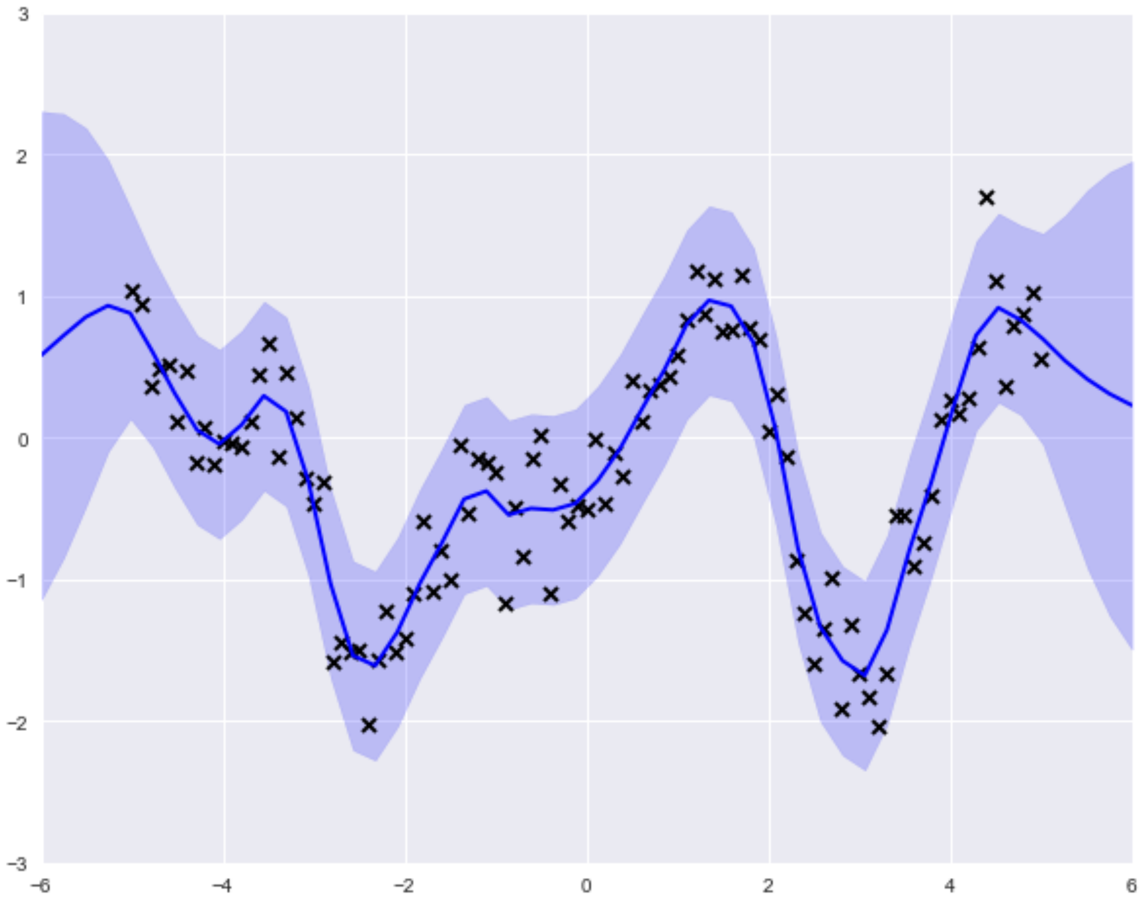
[Point Estimate]

$$p(\theta|D)$$

[Bayesian Estimate]

Example: $y = \theta_0 + \theta_1 \cdot x_1 + \ldots + \theta_k \cdot x_k$

Gaussian Processes



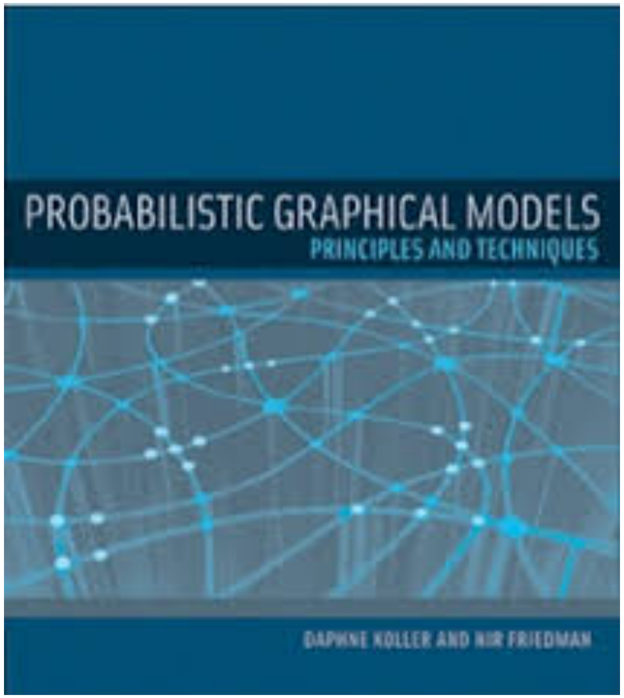
Missing Data

AMIDST TOOLBOX

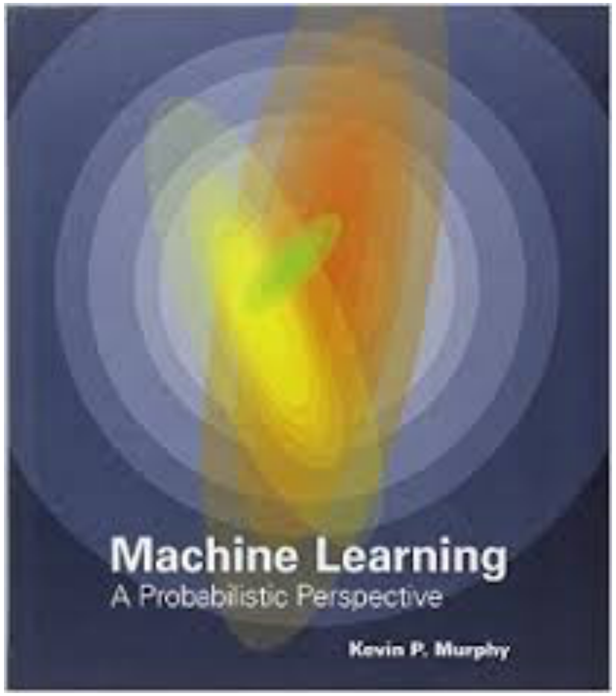
ID	S1_MaxTradeline	S2_BadLoans	S3_DeviceFirstSeen
841328	300	NA	11/16/2013
262927	500	0	10/1/2012
197305	750	0	NA
176415	NA	NA	NA
228986	0	3	NA
390908	800	NA	8/9/2013
846257	600	0	6/30/2012
254885	400	0	NA
833798	NA	0	3/9/2012
147660	900	2	NA

Probabilistic approach naturally deals with missing data

Everything is a random variable.



+



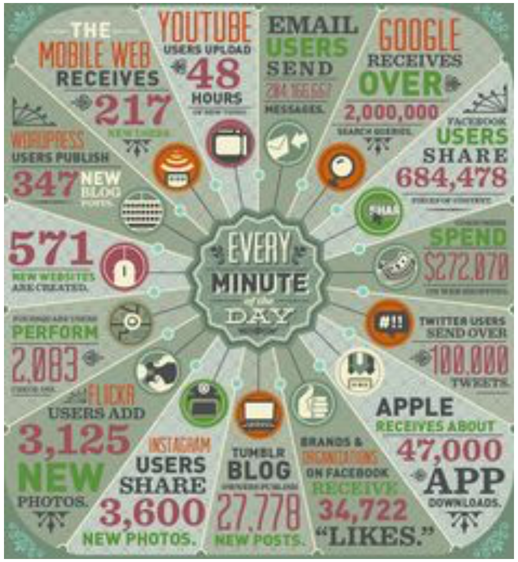
Probabilistic **Graphical Models**

Probabilistic Machine Learning

Data Streams

Update your models when new data is available.

DATA STREAMS



- Unbounded Flows of Data are generated daily:
  - Social Networks, sensors, network monitoring, finance, etc.
  - Continuous Model Updating.

Standard Approach

$$\{\theta : \theta \in \Theta\}$$

[Set of Models]

Data → $\boxed{\arg\min_{\theta \in \Theta} \ell(D_1, \theta)}$ → $\theta^\star$

[Optimal Model]

Loss Minimization

(Stochastic Gradient Descent)

STANDARD APPROACH

$$\{\theta : \theta \in \Theta\}$$

[Set of Models]

New Data → $\arg\min_{\theta \in \Theta} \ell(\{D_1 \cup D_2\}, \theta)$ → $\theta^\star$

[Optimal Model]

Loss Minimization

(Stochastic Gradient Descent)

AMIDST TOOLBOX

$$p(\theta|D_1)$$

[Previous Posterior]

New Data

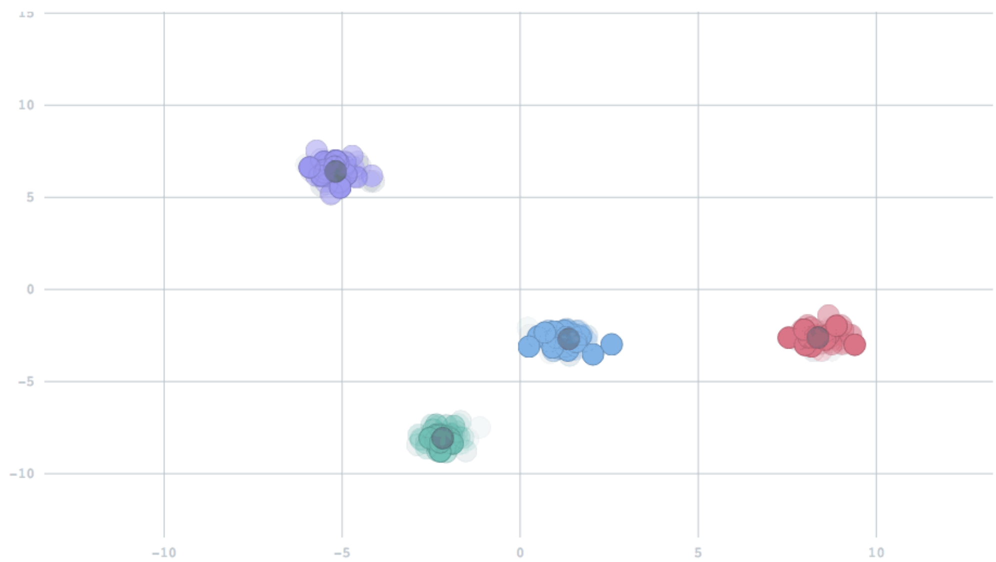
$$\boxed{\frac{p(D_2|\theta)p(\theta|D_1)}{p(D_2|D_1)}}$$

$$p(\theta|D_1, D_2)$$

[Updated Posterior]

Bayes Theorem

Freeman J. Introducing streaming k-means in Apache Spark 1.2.
https://databricks.com/blog/2015/01/28/introducing-streaming-k-means-in-spark-1-2.html



- Data may change from one time step to another.

DATA STREAMS

$$p(\theta|D_1)$$

[Previous Posterior]

New Data

$$\frac{p(D_2|\theta)p(D_1|\theta)^{\rho}p(\theta)}{p(D_1|D_2)}$$

$$\rho \in [0,1]$$

$$p(\theta|D_1, D_2)$$

[Updated Posterior]

Bayes Theorem
[Exponential Forgetting]

- Old-data is exponentially down-weighted.
  - Forgetting Mechanism. Focus on the present.

$$P(\theta|\mathbf{D})$$

Scalable Learning

Perform Bayesian inference on your probabilistic models with powerful approximate and scalable algorithms.

AMIDST TOOLBOX

$$p(\theta, H|D) = \frac{p(D|H,\theta)p(H|\theta)p(\theta)}{\int p(D|H,\theta)p(H|\theta)p(\theta)d\theta dH}$$

Highly Dimensional

Intractable Posterior

- Problem solving a highly multidimensional integral.
- Closed-form solution under very restrictive assumptions.
- Complex functional forms.

AMIDST TOOLBOX

$$\arg\min_{\lambda} KL(\overbrace{q(\theta, H|\lambda)}^{\text{Approximation}}||\overbrace{p(\theta, H|D)}^{\text{True Posterior}})$$

$$q(\boldsymbol{\theta}, \boldsymbol{H}) = \prod_{k=1}^{M} q(\theta_k) \prod_{i=1}^{N} \prod_{j=1}^{J} q(H_{i,j})$$

Variational Methods

- The inference problem is casted as an optimization problem.
- Deterministic approximation.

Hoffman, Matthew D., et al. "Stochastic variational inference." *Journal of Machine Learning Research* 14.1 (2013): 1303-1347.

$$\boxed{\ln p(D)} = \boxed{\mathcal{L}(\lambda)} + \boxed{KL(q(\theta, H|\lambda); p(\theta, H|D))}$$

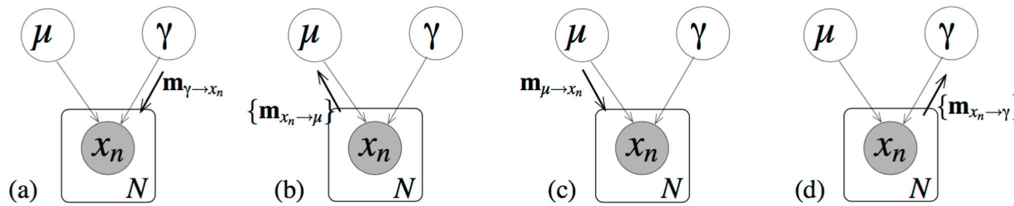
Constant Maximize Minimize

Optimization Problem

- The inference problem is casted as an optimization problem.
- Deterministic approximation.

Hoffman, Matthew D., et al. "Stochastic variational inference." *Journal of Machine Learning Research* 14.1 (2013): 1303-1347.

$$\frac{\partial \mathcal{L}}{\partial \lambda} =$$

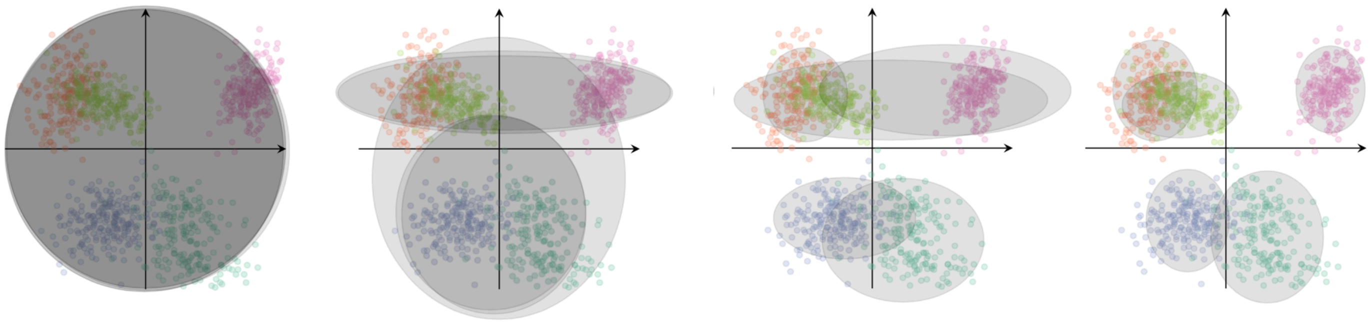


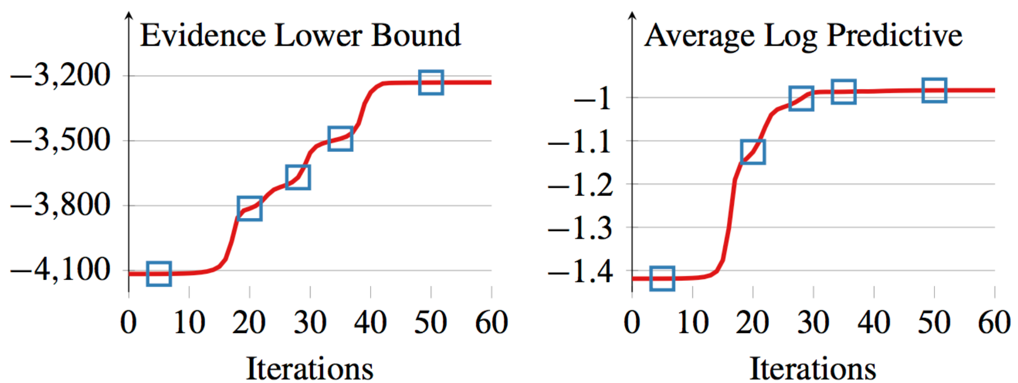
Variational Message Passing

- Automatic Gradient Computation
- Coordinate Ascent Algorithm

Winn, J., & Bishop, C. M. (2005). Variational message passing. *Journal of Machine Learning Research*, *6*(Apr), 661-694.

Variational Learning





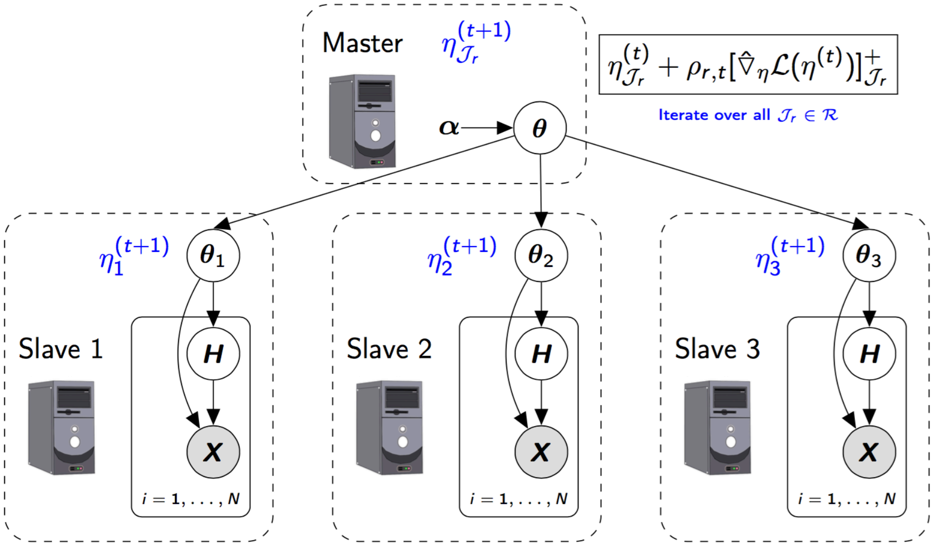
Winn, J., & Bishop, C. M. (2005). Variational message passing. *Journal of Machine Learning Research*, *6*(Apr), 661-694.

$$\lambda^{(t+1)} = \lambda^{(t)} + \rho \cdot N \cdot \frac{\partial \mathcal{L}(d_t, \lambda^{(t)})}{\partial \lambda}$$

Stochastic Gradient Ascent

- Estimate the gradient over a sub-sample of the data set

Hoffman, Matthew D., et al. "Stochastic variational inference." *Journal of Machine Learning Research* 14.1 (2013): 1303-1347.

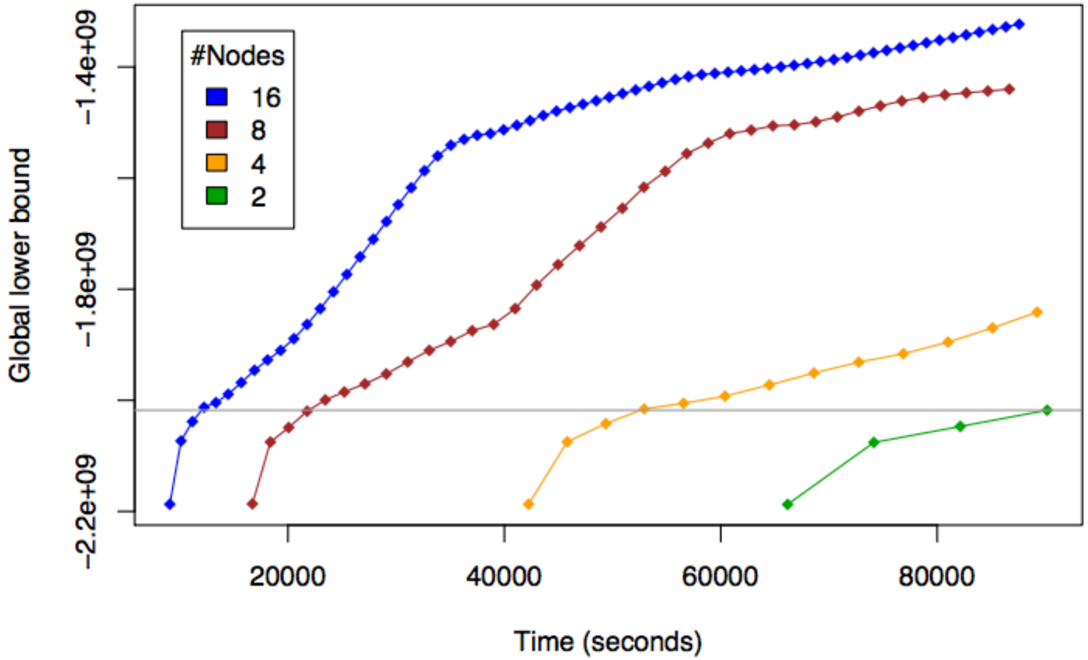


d-VMP Algorithm

A state-of-the-art distributed Variational Message Passing algorithm.

Masegosa, Andrés R., et al. "d-VMP: Distributed Variational Message Passing." *Proceedings of the Eighth International Conference on Probabilistic Graphical Models*. 2016.

AMIDST TOOLBOX



Masegosa, Andrés R., et al. "d-VMP: Distributed Variational Message Passing." *Proceedings of the Eighth International Conference on Probabilistic Graphical Models*. 2016.

One billion node probabilistic model

Experiment on a Flink cluster with 16 nodes on AWS.

Thanks for your attention

www www.amidsttoolbox.com

@ contact@amidsttoolbox.com

@AmidstToolbox

AMIDST TOOLBOX