AMiDST TOOLBOX

Bayesian Models of Data Streams

with Hierarchical Power Priors

Andres R. Masegosa(1), Thomas D. Nielsen(2), Helge Langseth(3), Dario Ramos-Lopez(1), Antonio Salmeron(1), Anders L. Madsen(2,4)

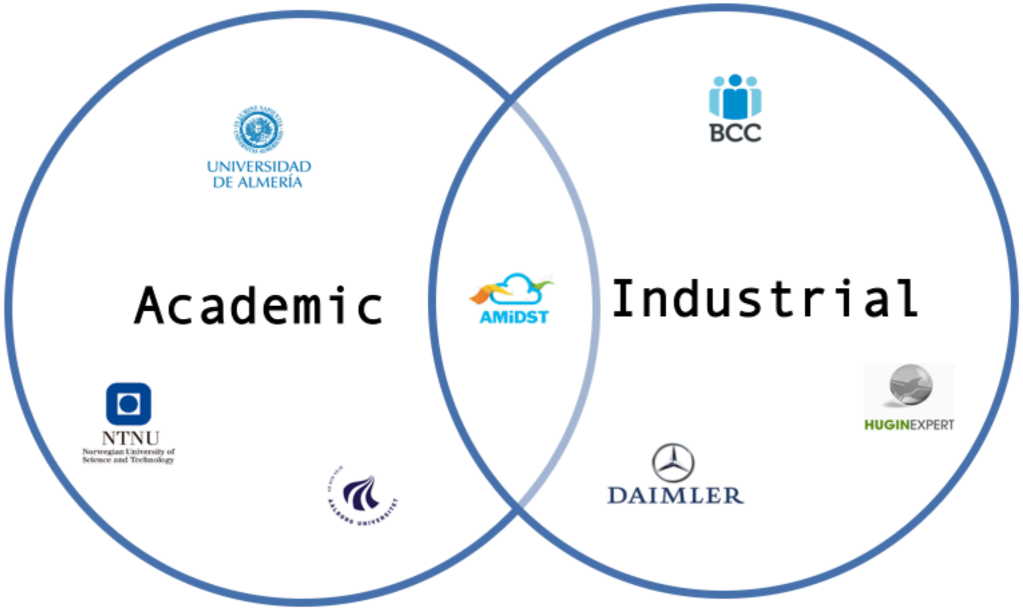
(1) University of Almeria
(2) University of Aalborg
(3) Norwegian University of Science and Technology
(4) Hugin Expert A/S

ICML2017, August 7th, Sydney

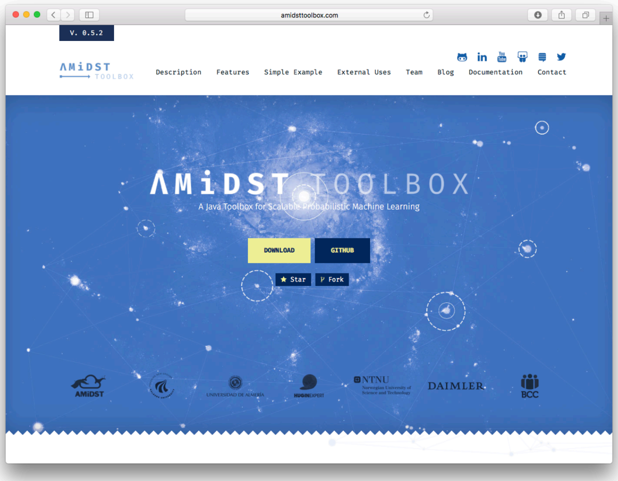
About us

The AMIDST consortium

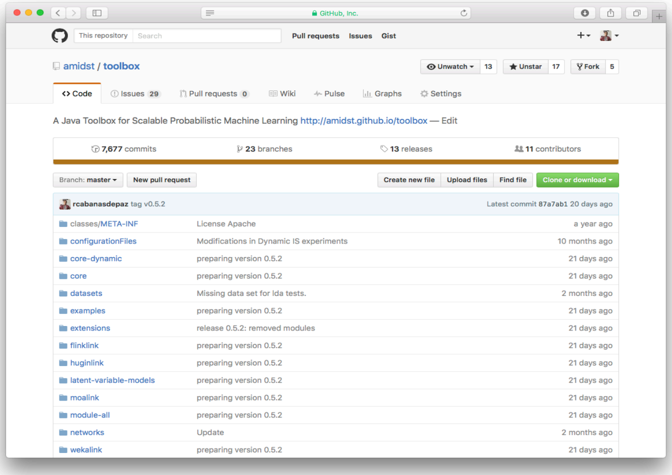
AMiDST TOOLBOX



A Java Toolbox for Scalable Probabilistic Machine Learning



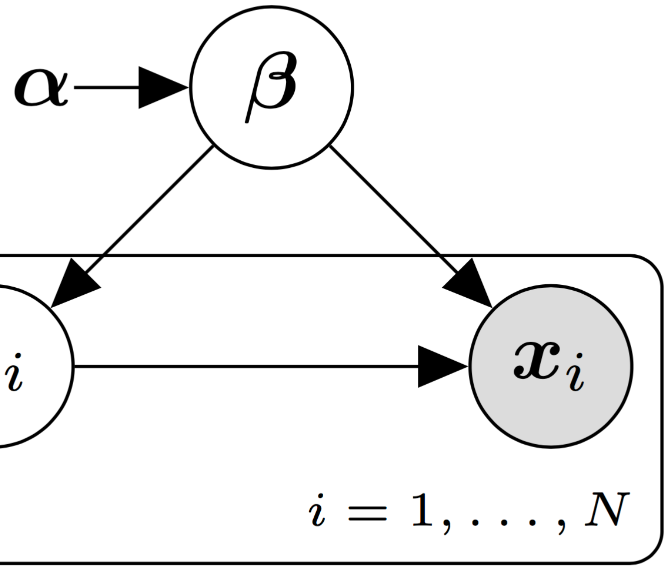
www.amidsttoolbox.com



github.com/amidst/toolbox

Apache License 2.0

The problem

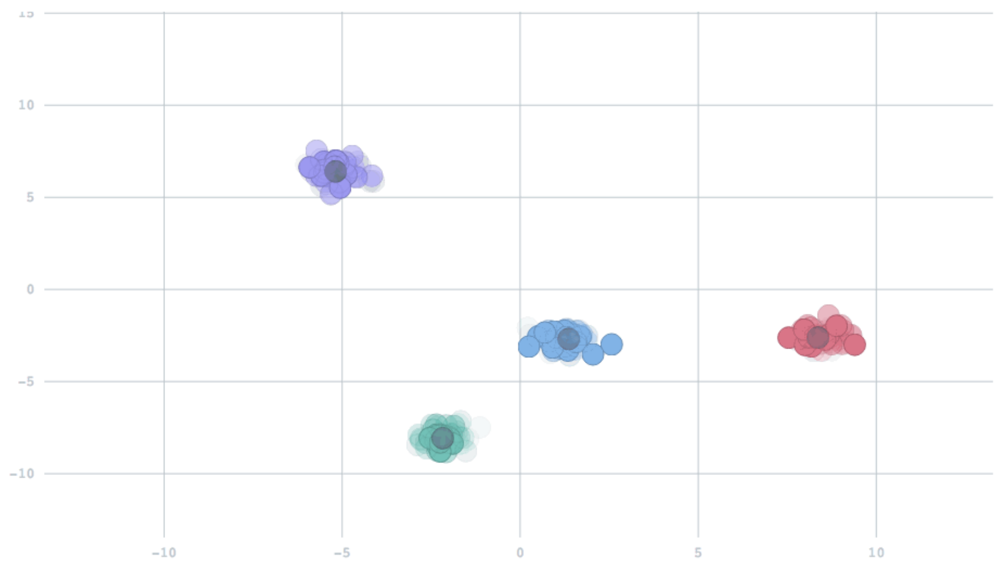


- **Variational Inference**
  - Latent Variable Models (LVMs).
  - Conjugate Exponential Family (CEF).

Winn & Bishop, 2005 Hoffman et al., 2013

AMiDST TOOLBOX

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

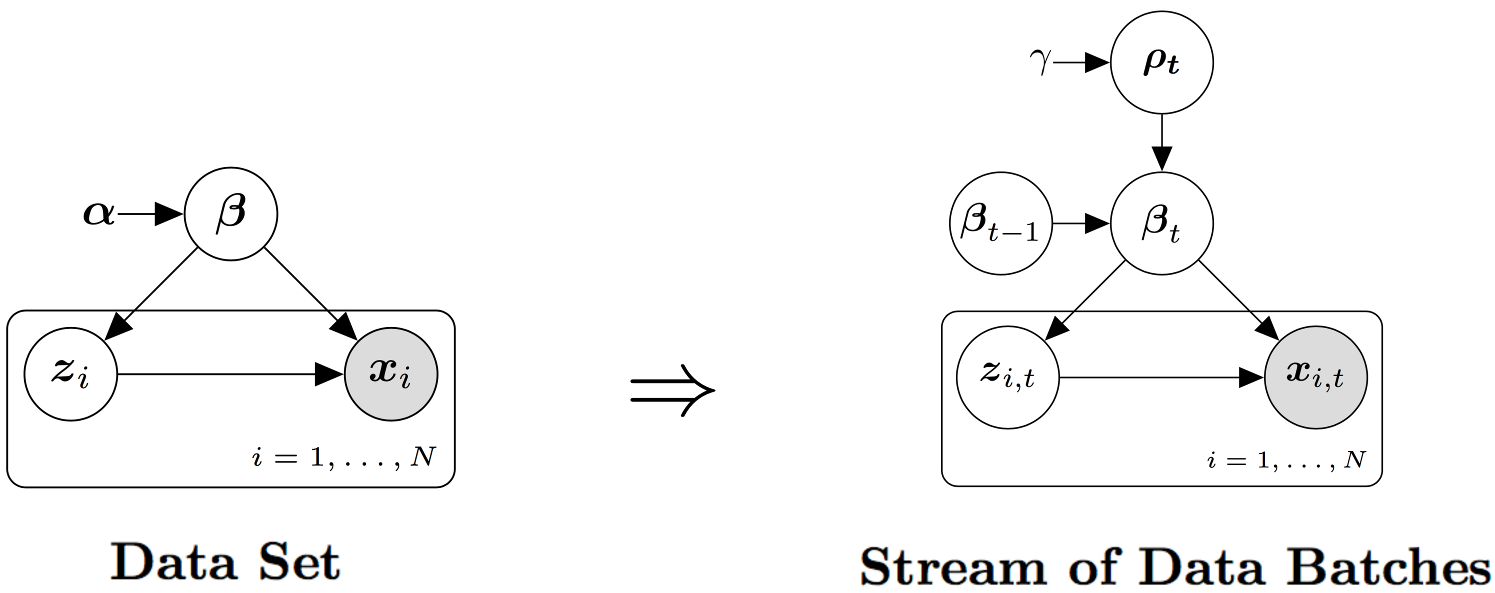


- **Learning from Data Streams**
  - Continuous Model Updating.
  - Bayesian posterior conditioned to non-finite data set.
  - Presence of Concept Drift (i.e. non i.i.d data).

Gama et al., 2014

Our proposal

Our Proposal

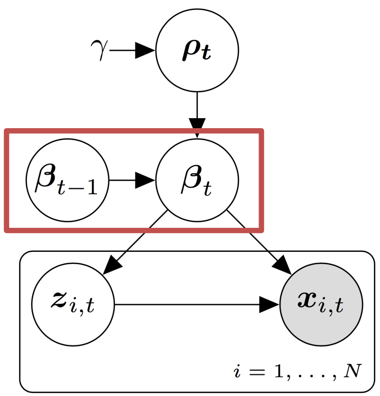


- **Out-of-the-box temporal extension.**
  - Global parameters β_t evolve over time.
  - Hierarchical prior modeling concept drift.
  - Closed-form Variational inference.

Implicit Transition Models

Kárnỳ (2014) and Özkan et al. (2013)

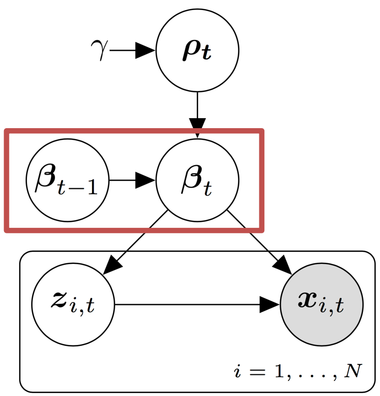
Explicit Transition Models



$$p(\boldsymbol{\beta}_t|\boldsymbol{x}_{1:t-1}) = \int \boxed{p(\boldsymbol{\beta}_t|\boldsymbol{\beta}_{t-1})} p(\boldsymbol{\beta}_{t-1}|\boldsymbol{x}_{1:t-1}) d\boldsymbol{\beta}_{t-1}$$

- **Explicit Transition Models**
  - Stationary transition model with requires domain knowledge.
  - Outside conjugate exponential family.

Explicit Transition Models

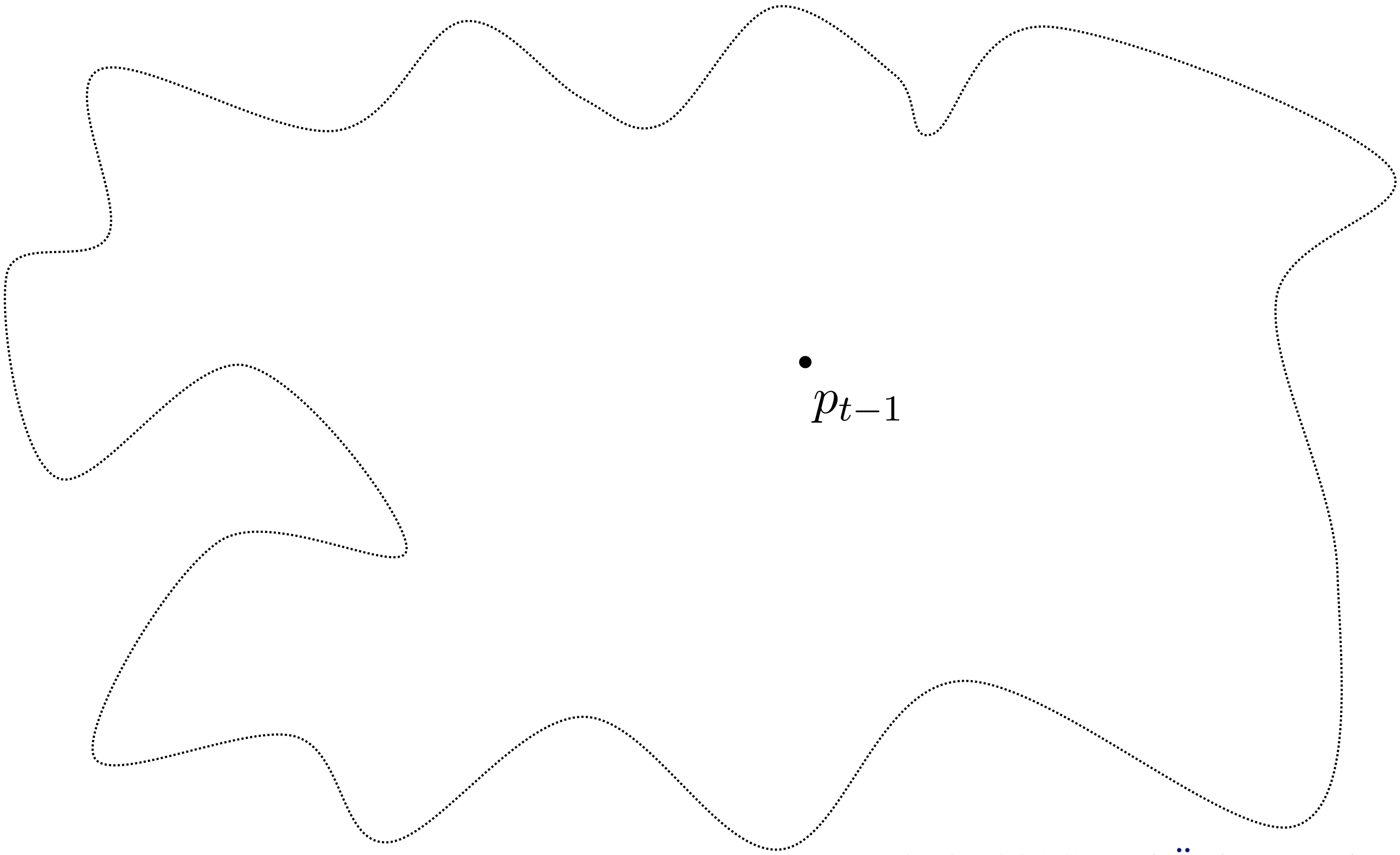


$$\overbrace{p(\boldsymbol{\beta}_t|\boldsymbol{x}_{1:t-1})}^{\hat{p}_t} = \int \boxed{p(\boldsymbol{\beta}_t|\boldsymbol{\beta}_{t-1})} \overbrace{p(\boldsymbol{\beta}_{t-1}|\boldsymbol{x}_{1:t-1})}^{p_{t-1}} d\boldsymbol{\beta}_{t-1}$$

- **Explicit Transition Models**
  - Stationary transition model with requires domain knowledge.
  - Outside conjugate exponential family.

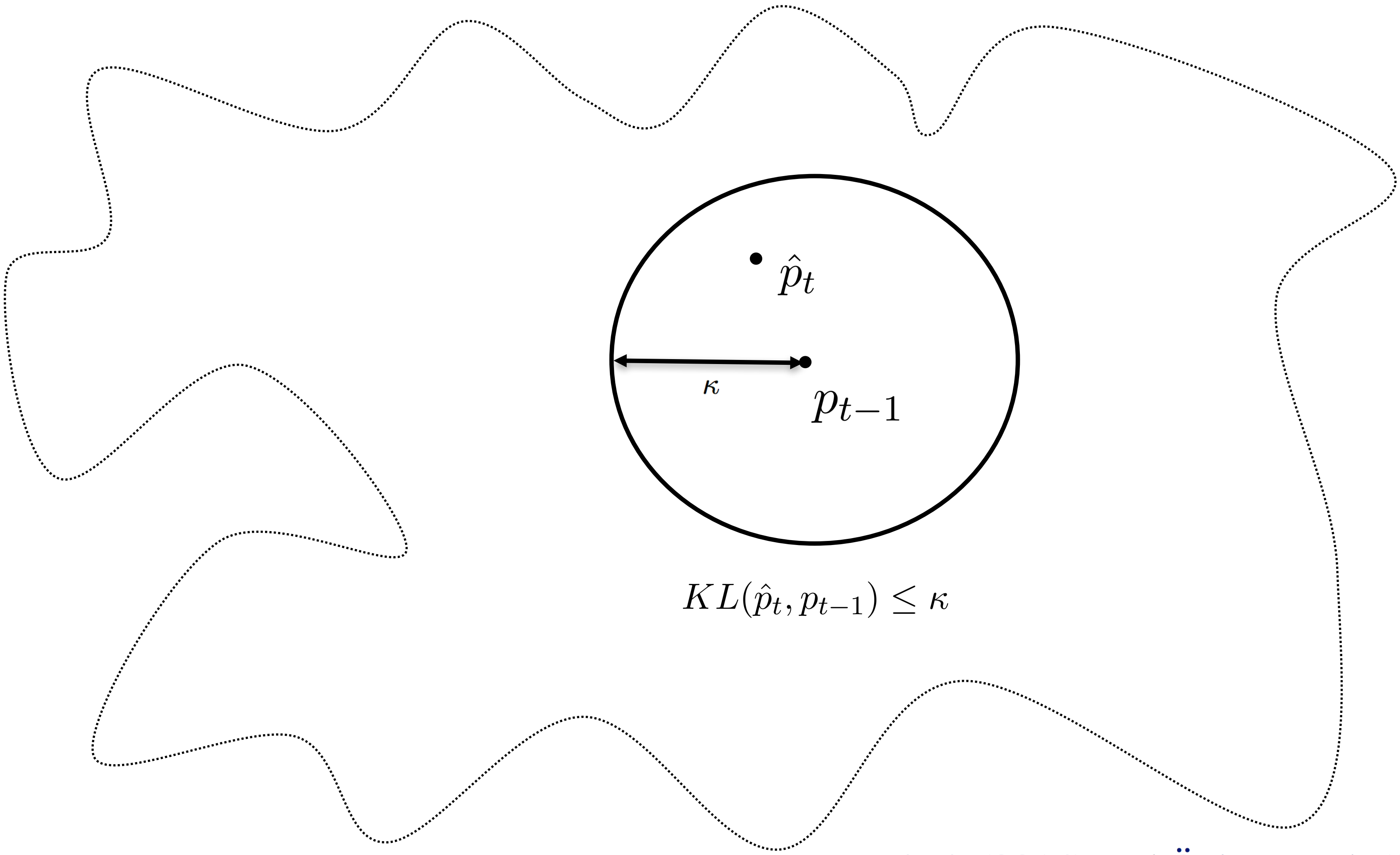
Implicit Transition Models

AMiDST TOOLBOX



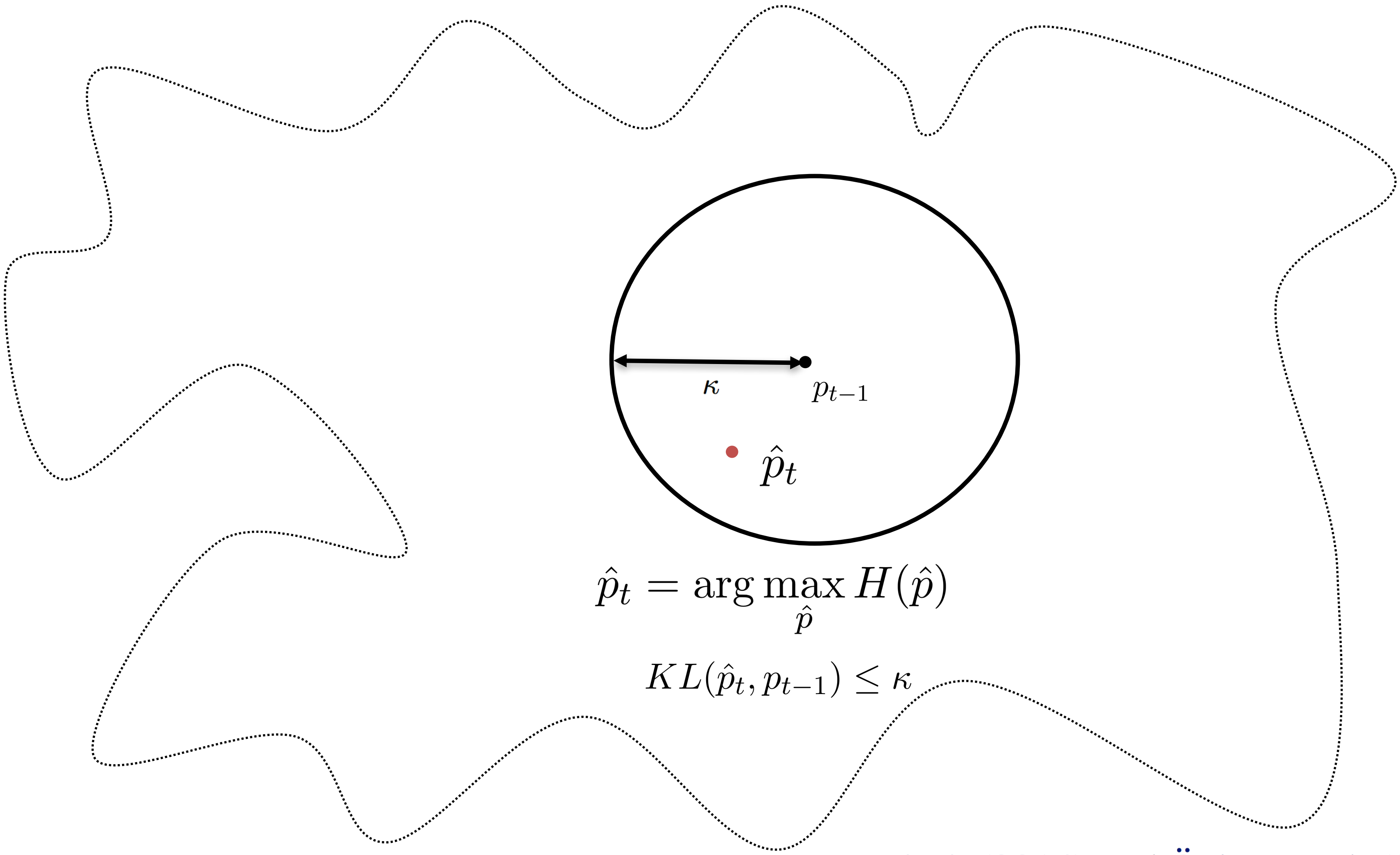
Kárnỳ (2014) and Özkan et al. (2013)

IMPLICIT TRANSITION MODELS



Kárnỳ (2014) and Özkan et al. (2013)

AMIDST TOOLBOX



$$\hat{p}_t = \arg\max_{\hat{p}} H(\hat{p})$$

$$KL(\hat{p}_t, p_{t-1}) \leq \kappa$$

Kárnỳ (2014) and Özkan et al. (2013)

$$\hat{\boldsymbol{\lambda}}_t = (1-\rho)\boldsymbol{\lambda}_u + \rho\boldsymbol{\lambda}_{t-1}$$

- **Closed-form solution for the Exponential Family**
  - λ natural parameter vector.
  - $\rho \in [0,1]$ is defined by the user.
  - $\rho = 1$ equals $\kappa = 0$ (i.e. maintain all the past data).
  - $\rho = 0$ equals $\kappa = \infty$ (i.e. completely forget past data).

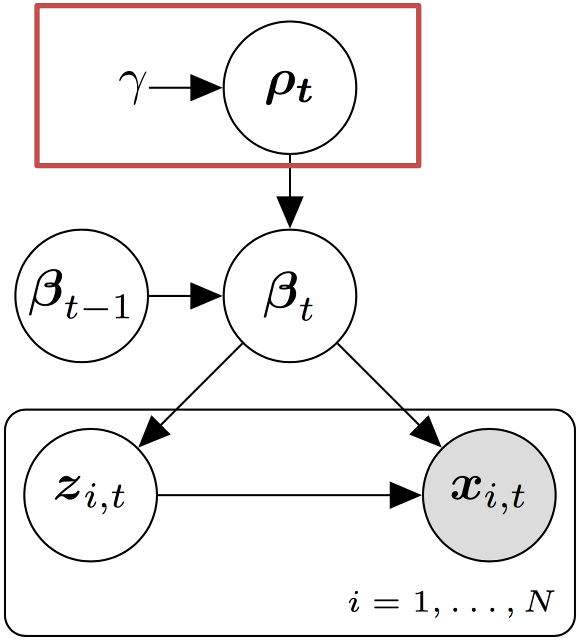
Kárnỳ (2014) and Özkan et al. (2013)

How to choose ρ?

- ρ defines the degree of forgetting.
- Optimal ρ is time dependent.

Hierarchical Power Priors

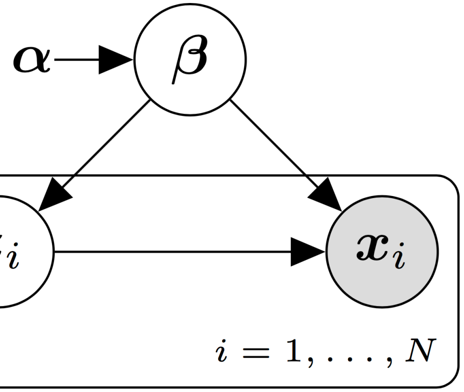
HIERARCHICAL POWER PRIORS



- $\rho_t \sim TrucantedExponential(\gamma)$, $\Omega(\rho_t) = [0, 1]$.
  - ρ_t close to 1 $\rightarrow$ No Drift at time t (i.e. $\beta_{t-1} \approx \beta_t$).
  - ρ_t close to 0 $\rightarrow$ Drift at time t (i.e. $\beta_{t-1} \not\approx \beta_t$).
- $p(\rho_t | \mathbf{x}_{1:t})$ tracks concept drift.

Variational Inference with HPPs

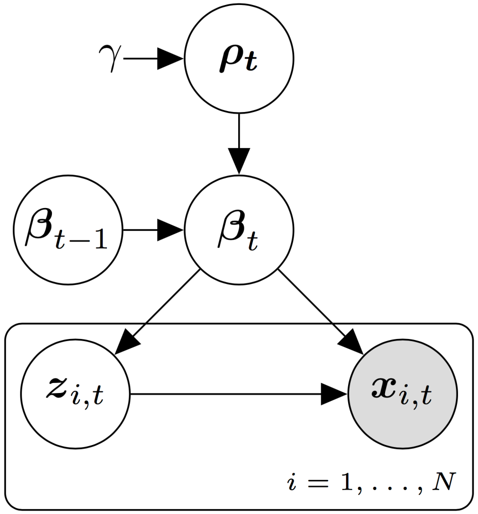
PREVIOUS KNOWLEDGE



$$q(\boldsymbol{\beta}, \boldsymbol{z}|\boldsymbol{\lambda}, \boldsymbol{\phi}) = \prod_{k=1}^{M} q(\beta_k|\lambda_k) \prod_{i=1}^{N} \prod_{j=1}^{J} q(z_{i,j}|\phi_{i,j})$$

- **Variational Inference in plain LVMs**
  - $(\boldsymbol{\lambda}^\star, \boldsymbol{\phi}^\star) = \arg\max_{\boldsymbol{\lambda}, \boldsymbol{\phi}} \mathcal{L}(\boldsymbol{\lambda}, \boldsymbol{\phi}|\mathbf{x}, \boldsymbol{\alpha})$
  - Closed-form gradients for CEF models.

Winn & Bishop, 2005 Hoffman et al., 2013



$$q(\boldsymbol{\beta}_t, \boldsymbol{z}_t, \rho_t | \boldsymbol{\lambda}_t, \boldsymbol{\phi}_t, \omega_t)$$

- **Variational Inference in temporal LVMs**
  - $(\boldsymbol{\lambda}_t^\star, \boldsymbol{\phi}_t^\star, \omega_t^\star) = \arg\max_{\boldsymbol{\lambda}_t, \boldsymbol{\phi}_t, \omega_t} \mathcal{L}_{HPP}(\boldsymbol{\lambda}_t, \boldsymbol{\phi}_t, \omega_t | \mathbf{x}_t, \boldsymbol{\lambda}_{t-1})$
  - **No closed-form gradients.**

AMiDST TOOLBOX

$$\mathcal{L}_{HPP} \geq \hat{\mathcal{L}}_{HPP}$$

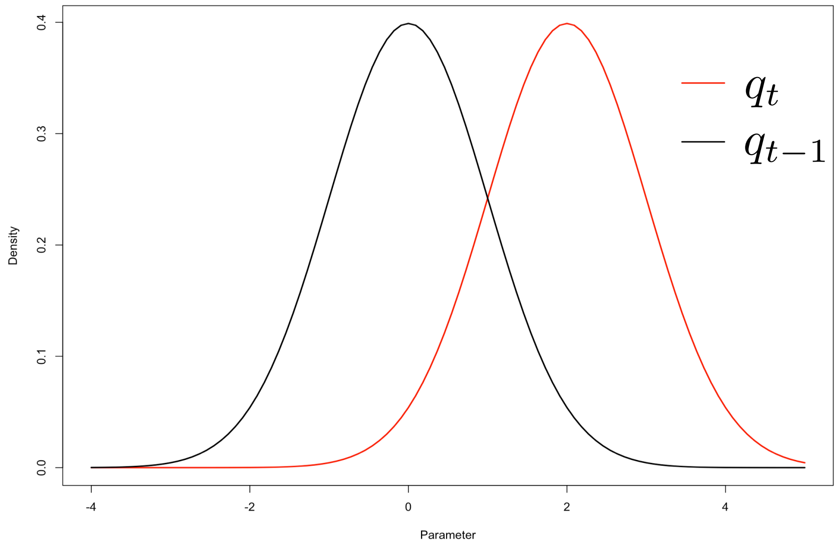
- **A double-lower bound**
  - $\frac{\partial \hat{\mathcal{L}}_{HPP}}{\partial \boldsymbol{\lambda}_t} = \frac{\partial \mathcal{L}}{\partial \boldsymbol{\lambda}}$
  - $\frac{\partial \hat{\mathcal{L}}_{HPP}}{\partial \boldsymbol{\phi}_t} = \frac{\partial \mathcal{L}}{\partial \boldsymbol{\phi}}$

with $\boldsymbol{\alpha} = (1 - E_q[\rho_t])\boldsymbol{\lambda}_u + E_q[\rho_t]\boldsymbol{\lambda}_{t-1}$

- **Closed-form gradient**
  - $\frac{\partial \hat{\mathcal{L}}_{HPP}}{\partial \omega_t} = KL(q_t, p_u) - KL(q_t, q_{t-1}) + \gamma - \omega_t.$
  - A measure of concept drift.

AMiDST TOOLBOX

Drift

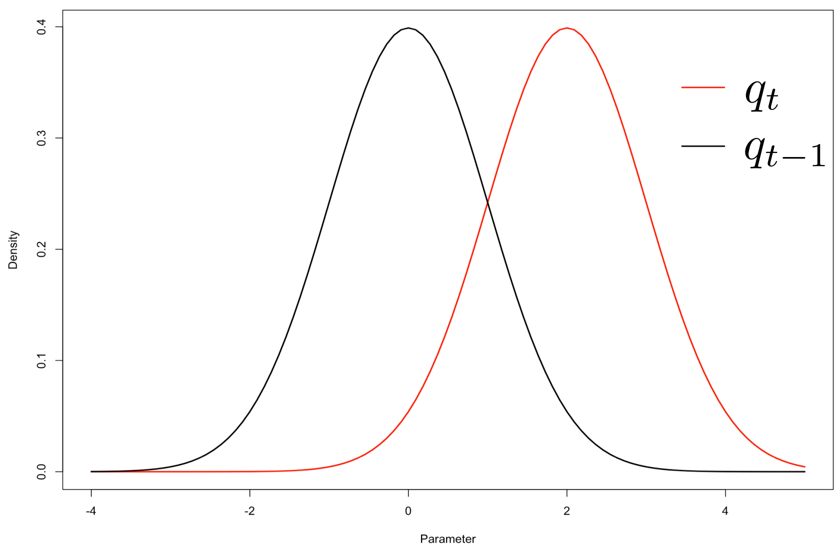


$$KL(q_t, p_u) + \gamma < KL(q_t, q_{t-1})$$

- **Closed-form gradient**
  - $\frac{\partial \hat{\mathcal{L}}_{HPP}}{\partial \omega_t} = KL(q_t, p_u) - KL(q_t, q_{t-1}) + \gamma - \omega_t.$
  - A measure of concept drift.

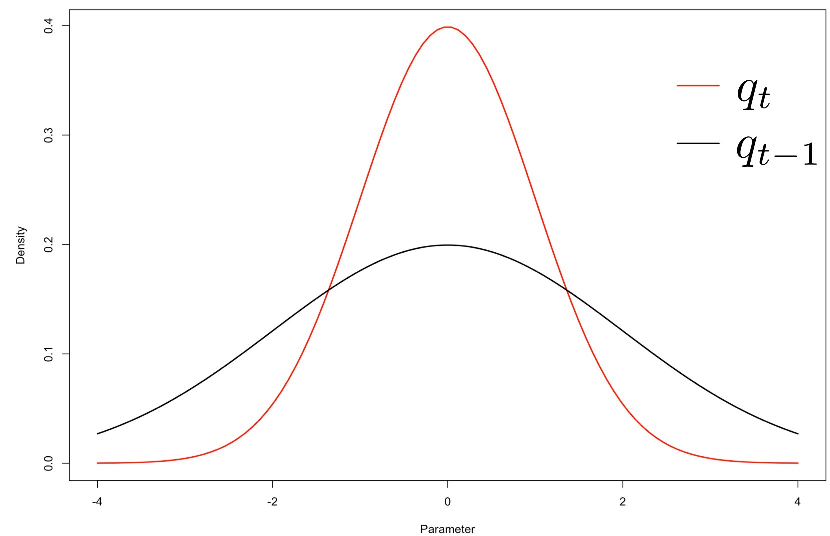
MEASURING CONCEPT DRIFT

Drift



$$KL(q_t, p_u) + \gamma < KL(q_t, q_{t-1})$$

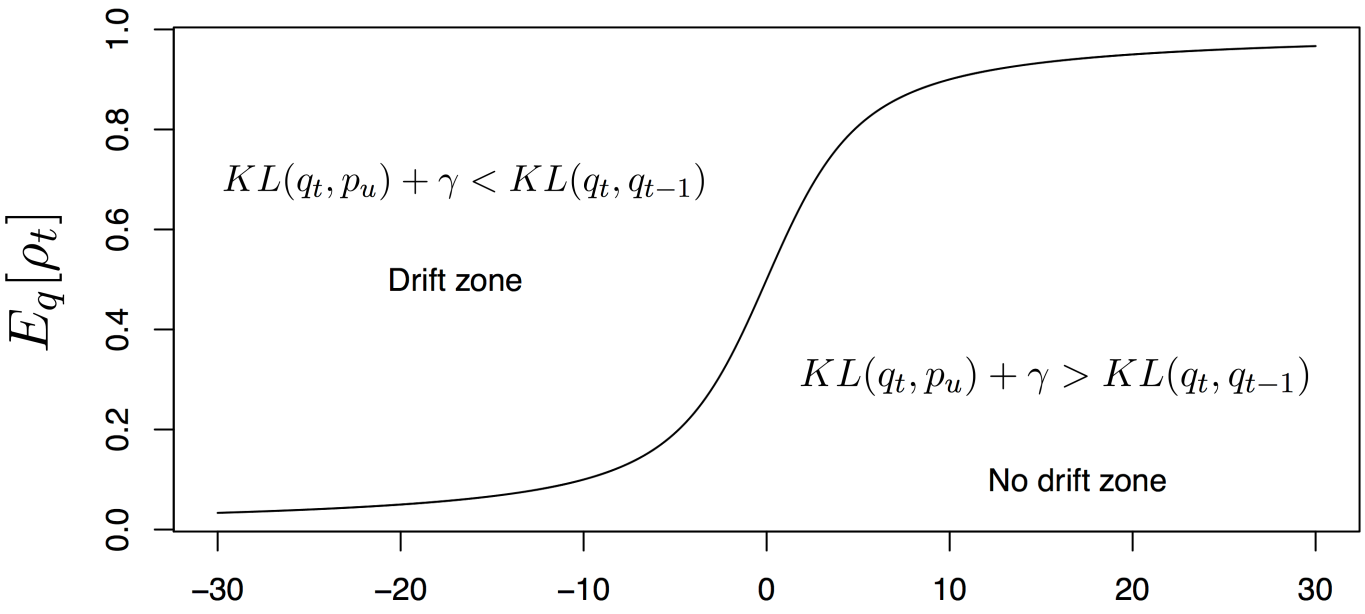
No Drift



$$KL(q_t, p_u) + \gamma > KL(q_t, q_{t-1})$$

- **Closed-form gradient**
  - $\frac{\partial \hat{\mathcal{L}}_{HPP}}{\partial \omega_t} = KL(q_t, p_u) - KL(q_t, q_{t-1}) + \gamma - \omega_t.$
  - A measure of concept drift.

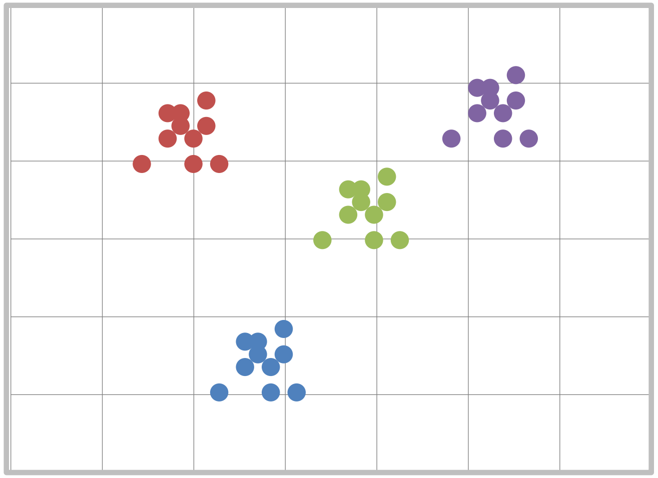
Measuring concept drift



- **Closed-form gradient**
  - $\frac{\partial \hat{\mathcal{L}}_{HPP}}{\partial \omega_t} = KL(q_t, p_u) - KL(q_t, q_{t-1}) + \gamma - \omega_t.$
  - A measure of concept drift.

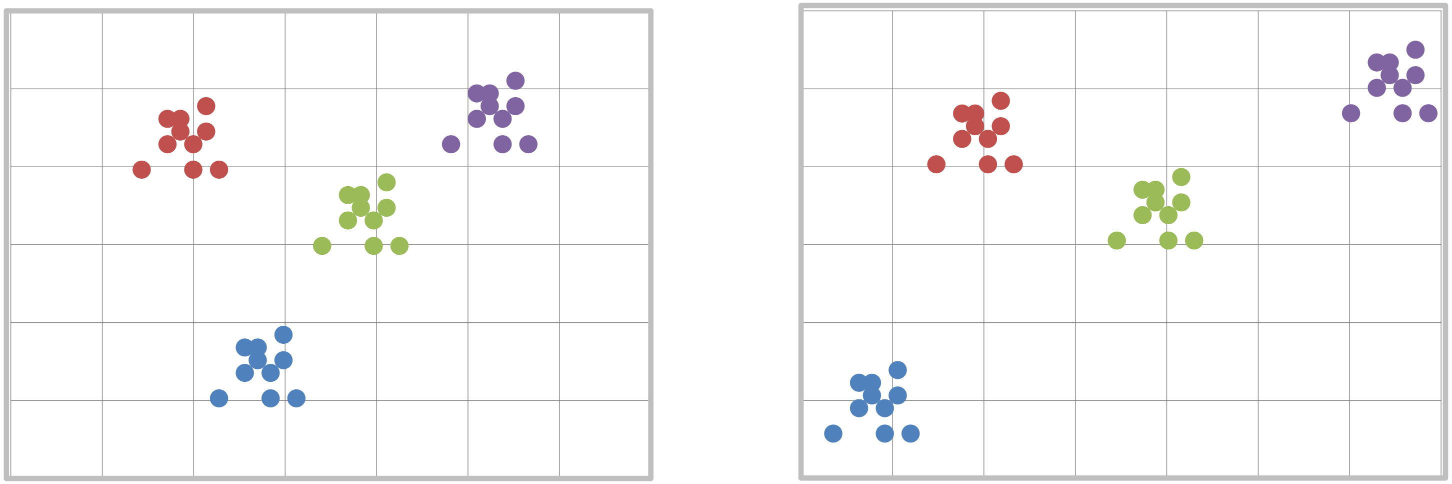
What if only part of the data drifts?

Multiple HPPs



$$\mathbf{x}_{t-1}$$

MULTIPLE HPPS

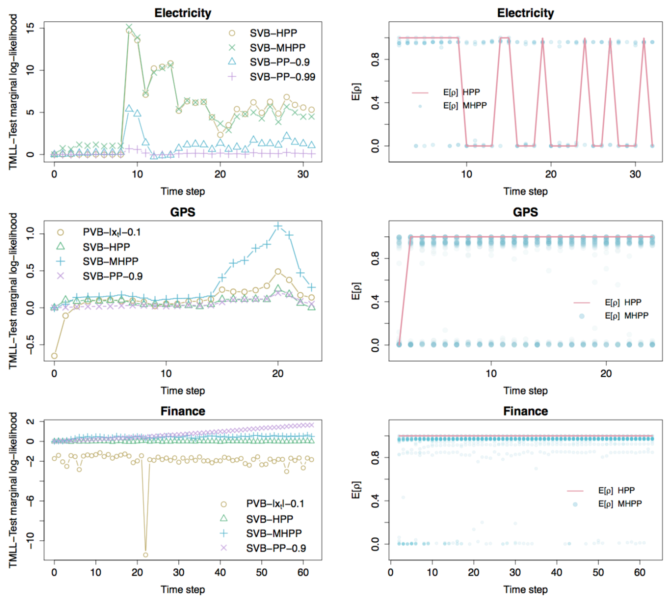


$\mathbf{x}_{t-1}$ $\mathbf{x}_t$

- **Multiple HPPs**
  - Place independent $\rho_{k,t}$ for each parameter of the model.
  - Closed-form Variational inference.

Experimental Evaluation

Experimental Evaluation

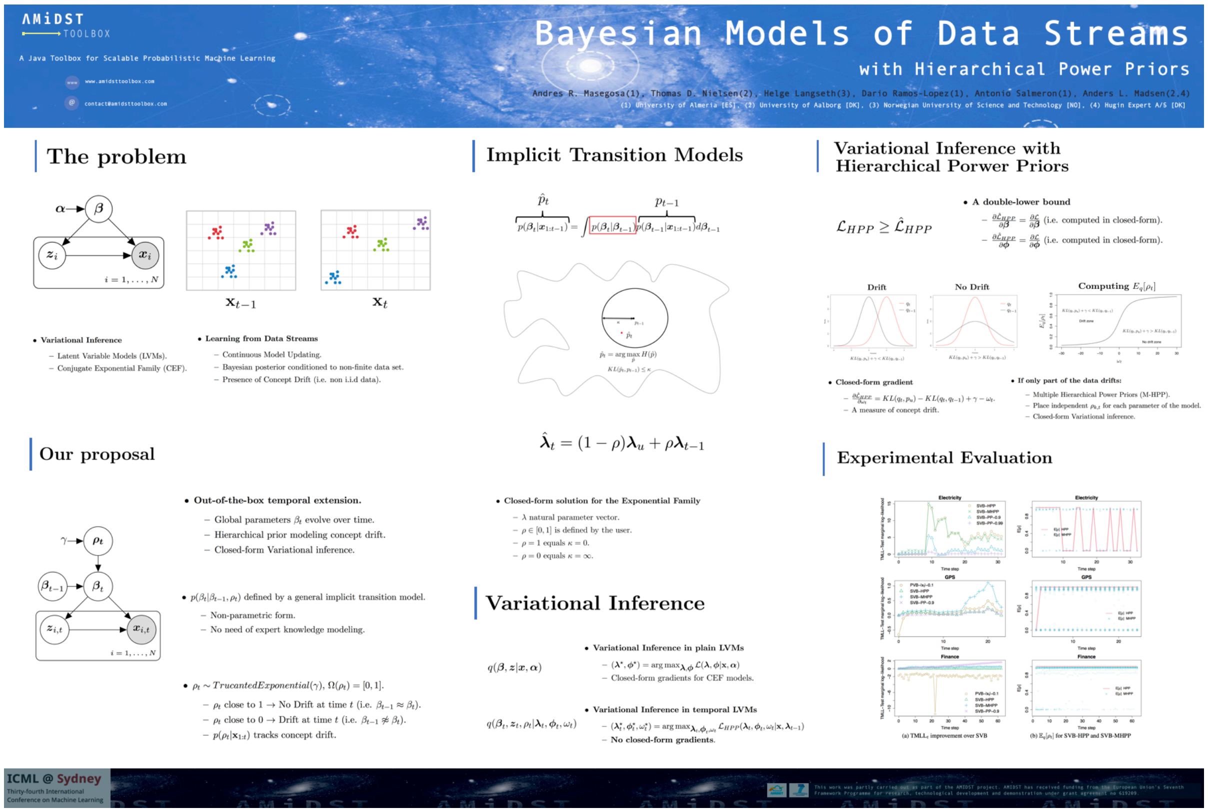


(a) TMLL_t improvement over SVB (b) $\mathbb{E}_q[\rho_t]$ for SVB-HPP and SVB-MHPP

- **Summary of the evaluation:**
  - M-HPP is the most robust approach.
  - Adaptive forgetting mechanisms are usually needed.
  - Concept drift usually affects only a part of the model.

COME TO MY POSTER @ GALLERY #37

AMIDST TOOLBOX



Thanks for your attention

www www.amidsttoolbox.com

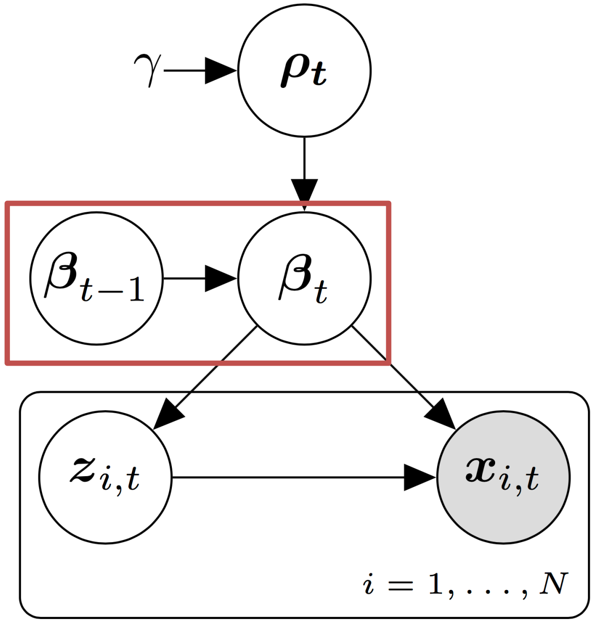
@ contact@amidsttoolbox.com

@AmidstToolbox

AMiDST TOOLBOX

Implicit Transition Models

AMiDST TOOLBOX



- $p(\beta_t | \beta_{t-1}, \rho_t)$ defined by a general implicit transition model.
  - Non-parametric form.
  - No need of expert knowledge modeling.