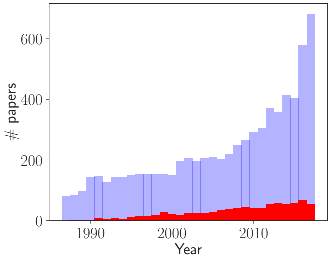
# Bayesian machine learning

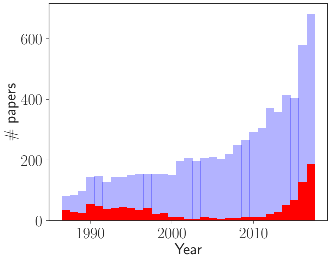
Rémi Bardenet & (later) Julyan Arbel

CNRS & CRIStAL, Univ. Lille, France

# Trends in abstracts




(a) "Bayesian" @NeurIPS




(b) "Neural" @NeurIPS

- The Bayesian deep learning workshop is increasingly popular (accepted papers: 39 in 2016, 68 in 2017 and 136 in 2018).

# We can extract topics

| |
|---|
| model models data process latent Bayesian Dirichlet hierarchical nonparametric inference |
| features learn problem different knowledge learning image object example examples |
| method neural Bayesian using linear state based kernel approach model |
| belief propagation nodes local tree posterior node nbsp given algorithm |
| learning data Bayesian model training classification performance selection prediction sets |
| inference Monte Carlo Markov sampling variational time algorithm MCMC approximate |
| function optimization algorithm optimal learning problem gradient methods bounds state |
| learning networks variables structure network Bayesian EM paper distribution algorithm |
| Bayesian gaussian prior regression non estimation likelihood sparse parameters matrix |
| model information Bayesian human visual task probability sensory prior concept |

**Figure:** Topics extracted from NeurIPS and JMLR abstracts using variational inference for latent Dirichlet allocation.

# Goals of the course

By the end of the course, the students should

- have a high-level view of the main approaches to making decisions under uncertainty.
- be able to detect when being Bayesian helps and why.
- be able to design and run a Bayesian ML pipeline for standard supervised or unsupervised learning.
- have a global view of the current limitations of Bayesian approaches and the research landscape.
- be able to understand the abstract of most Bayesian ML papers.

# High-level outline

1. Principles of decision theory [3, 2]
2. Formalizing a problem in a Bayesian way [3, 4]
3. Bayesian computation: MCMC and variational methods [5, 1]
4. Bayesian nonparametrics
5. Bayesian deep learning

**Course webpage**

Check out `github.com/rbardenet/bml-course`

**There will be four practical sessions**

- Bring your laptops.
- We'll start with Python. Make sure you install
  - Anaconda with Python 3.5+
  - pymc3 with dependencies (including theano)
  - run a few examples from pymc3's docs to check.

# High-level outline

1. Principles of decision theory [3, 2]
2. Formalizing a problem in a Bayesian way [3, 4]
3. Bayesian computation: MCMC and variational methods [5, 1]
4. Bayesian nonparametrics
5. Bayesian deep learning

**Course webpage**

Check out `github.com/rbardenet/bml-course`

**There will be four practical sessions**

- Bring your laptops.
- We'll start with Python. Make sure you install
  - Anaconda with Python 3.5+
  - pymc3 with dependencies (including theano)
  - run a few examples from pymc3's docs to check.

# If you're looking for a PhD on Monte Carlo methods

We are hiring, check out

`rbardenet.github.io/pdf/phd-proposal.pdf`

and write me an email quickly with vita and transcripts.

European Research Council

Established by the European Commission

## Outline

### Waldian decisions and objective Bayes

### Subjective Bayes

### More examples of Bayesian decision problems in ML

### Computation: MCMC

### Metropolis-Hastings

### Computation: variational Bayes

# Outline

# Outline

Waldian decisions and objective Bayes

Subjective Bayes

**More examples of Bayesian decision problems in ML**

Computation: MCMC

Metropolis-Hastings

Computation: variational Bayes

# Outline

# Outline

# Metropolis-Hastings

$\mathrm{MH}\big(\pi(x),\ q(y|x),\ x_0,\ N_{\text{iter}}\big)$

**for** $k \leftarrow 1$ **to** $N_{\mathrm{iter}}$

$\quad x \leftarrow x_{k-1}$

$\quad y \sim q(.|x),\ u \sim \mathcal{U}_{(0,1)},$

$\quad \alpha = \frac{\pi(y)}{\pi(x)} \frac{q(x|y)}{q(y|x)}$

$\quad$ **if** $u < \alpha$

$\quad\quad x_k \leftarrow y \qquad \triangleright$ *Accept*

$\quad$ **else** $x_k \leftarrow x \qquad \triangleright$ *Reject*

**return** $(x_k)_{k=1,\dots,N_{\mathrm{iter}}}$

- Under weak assumptions, see Chapter 7 of [**DoMoSt14**],

$$\sqrt{N_{\text{iter}}}\left[\frac{1}{N_{\text{iter}}}\sum_{k=0}^{N_{\text{iter}}} h(x_k) - \int h(x)\,\pi(x)dx\right] \to \mathcal{N}(0, \sigma^2_{\text{lim}}(h)),$$

# Metropolis-Hastings

$\mathrm{MH}\big(\pi(x),\ q(y|x),\ x_0,\ N_{\text{iter}}\big)$

**for** $k \leftarrow 1$ **to** $N_{\mathrm{iter}}$

$\quad x \leftarrow x_{k-1}$

$\quad y \sim q(.|x),\ u \sim \mathcal{U}_{(0,1)},$

$\quad \alpha = \frac{\pi(y)}{\pi(x)} \frac{q(x|y)}{q(y|x)}$

$\quad$ **if** $u < \alpha$

$\quad\quad x_k \leftarrow y \qquad \triangleright$ *Accept*

$\quad$ **else** $x_k \leftarrow x \qquad \triangleright$ *Reject*

**return** $(x_k)_{k=1,\dots,N_{\mathrm{iter}}}$

- Under weak assumptions, see Chapter 7 of [**DoMoSt14**],

$$\sqrt{N_{\text{iter}}}\left[\frac{1}{N_{\text{iter}}}\sum_{k=0}^{N_{\text{iter}}} h(x_k) - \int h(x)\,\pi(x)dx\right] \to \mathcal{N}(0, \sigma^2_{\text{lim}}(h)),$$

## Metropolis-Hastings

$\mathrm{MH}\big(\pi(x),\ q(y|x),\ x_0,\ N_{\text{iter}}\big)$

```
for k ← 1 to N_iter
    x ← x_{k−1}
    y ∼ q(.|x),  u ∼ U_(0,1),
    α = π(y)/π(x) · q(x|y)/q(y|x)
    if u < α
        x_k ← y          ▷ Accept
    else x_k ← x         ▷ Reject
return (x_k)_{k=1,...,N_iter}
```

- Under weak assumptions, see Chapter 7 of [**DoMoSt14**],

$$\sqrt{N_{\text{iter}}}\left[\frac{1}{N_{\text{iter}}}\sum_{k=0}^{N_{\text{iter}}} h(x_k) - \int h(x)\,\pi(x)dx\right] \to \mathcal{N}(0, \sigma^2_{\text{lim}}(h)),$$

## Metropolis-Hastings

$\mathrm{MH}\big(\pi(x),\ q(y|x),\ x_0,\ N_{\text{iter}}\big)$

**for** $k \leftarrow 1$ **to** $N_{\mathrm{iter}}$

  $x \leftarrow x_{k-1}$

  $y \sim q(.|x),\ u \sim \mathcal{U}_{(0,1)},$

  $\alpha = \frac{\pi(y)}{\pi(x)} \frac{q(x|y)}{q(y|x)}$

  **if** $u < \alpha$

    $x_k \leftarrow y$  ▷ *Accept*

  **else** $x_k \leftarrow x$  ▷ *Reject*

**return** $(x_k)_{k=1,\dots,N_{\mathrm{iter}}}$

- Under weak assumptions, see Chapter 7 of [**DoMoSt14**],

$$\sqrt{N_{\text{iter}}}\left[\frac{1}{N_{\text{iter}}}\sum_{k=0}^{N_{\text{iter}}} h(x_k) - \int h(x)\,\pi(x)dx\right] \to \mathcal{N}(0, \sigma^2_{\text{lim}}(h)),$$

## Metropolis-Hastings

$\mathrm{MH}\big(\pi(x),\ q(y|x),\ x_0,\ N_{\mathrm{iter}}\big)$

**for** $k \leftarrow 1$ **to** $N_{\mathrm{iter}}$
    $x \leftarrow x_{k-1}$
    $y \sim q(.|x),\ u \sim \mathcal{U}_{(0,1)},$
    $\alpha = \frac{\pi(y)}{\pi(x)} \frac{q(x|y)}{q(y|x)}$
    **if** $u < \alpha$
        $x_k \leftarrow y$  ▷ *Accept*
    **else** $x_k \leftarrow x$  ▷ *Reject*
**return** $(x_k)_{k=1,\dots,N_{\mathrm{iter}}}$

- Under weak assumptions, see Chapter 7 of [**DoMoSt14**],

$$\sqrt{N_{\mathrm{iter}}}\left[\frac{1}{N_{\mathrm{iter}}}\sum_{k=0}^{N_{\mathrm{iter}}} h(x_k) - \int h(x)\,\pi(x)dx\right] \to \mathcal{N}(0, \sigma^2_{\mathrm{lim}}(h)),$$

## Outline

Waldian decisions and objective Bayes

Subjective Bayes

More examples of Bayesian decision problems in ML

Computation: MCMC

Metropolis-Hastings

**Computation: variational Bayes**

# References I


[1] D. M. Blei, A. Kucukelbir, and J. D. McAuliffe. "Variational inference: A review for statisticians". In: *Journal of the American Statistical Association* 112.518 (2017), pp. 859–877.

[2] D. Kreps. *Notes on the Theory of Choice*. Westview press, 1988.

[3] G. Parmigiani and L. Inoue. *Decision theory: principles and approaches*. Vol. 812. John Wiley & Sons, 2009.

[4] C. P. Robert. *The Bayesian choice: from decision-theoretic foundations to computational implementation*. Springer Science & Business Media, 2007.

[5] C. P. Robert and G. Casella. *Monte Carlo statistical methods*. Springer, 2004.