

Être bayésien quand on a trop de données

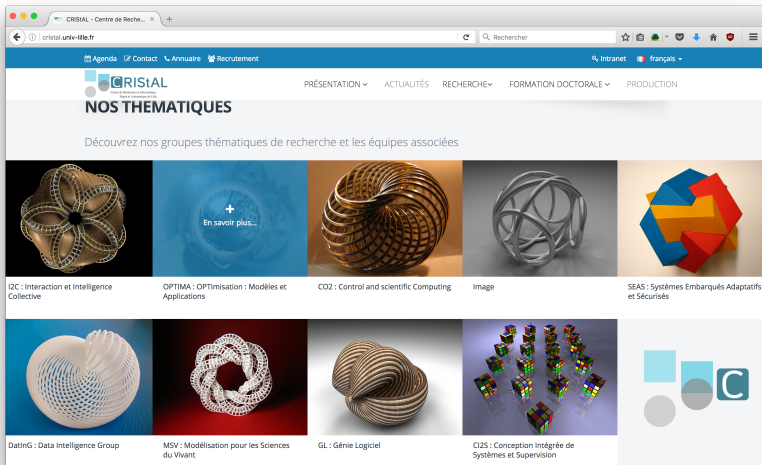
Précédé d'une introduction au mille-feuille CRISAL

Rémi Bardenet¹

¹CNRS & CRISAL, Univ. Lille, France



The screenshot shows the homepage of the University of Lille website. At the top, there is a navigation bar with the university logo and name, a search bar, and an 'Intranet' link. Below this is a main menu with categories: 'DU PROJET À LA FUSION', 'FORMATION', 'RECHERCHE', 'INTERNATIONAL', and 'VIE DES CAMPUS'. The main content area features a large graphic with the text 'BONNE ANNÉE' and '2017 ENSEMBLE' in the center, surrounded by names of Lille districts: Lille, Tourcoing, Roubaix, Willemeux, Villeneuve d'Ascq, Loos, Ronchin, and Cambrai. To the right, there are two promotional banners: one for 'L'UNIVERSITÉ DE LILLE. PARTENAIRE DE O21 le Monde S'ORIENTER AU 21^È SIÈCLE' and another for 'DEVENIR ACTEUR DU MONDE'. Below the main graphic, there are four news items: 'Prix de l'innovation 2016 : l'équipe lilloise TBBBoost lauréate', 'Fertilité masculine : SPOJ lauréate des partenariats régionaux d'innovation', 'Force Awards 2016 : 2 trophées pour la société SPOJ', and 'Toute l'actualité'. At the bottom, there is a section titled 'DU PROJET À LA FUSION' with three sub-sections: 'Stratégie politique' (Les trois universités lilloises se mobilisent), 'Les membres' (Huit écoles publiques de la région), and 'Plan campus' (L'Université de Lille porte un des 12).



- ▶ 222 permanents dont 22 CNRS et 27 Inria.

GT - CRIStAL

https://www.cristal.univ-lille.fr/trubrique265?id=4

Agenda Contact Annuaire Recrutement

Intranet français

PRÉSENTATION ACTUALITÉS RECHERCHE FORMATION DOCTORALE PRODUCTION

DatIng : Data Intelligence Group

Responsable: Philippe Preux

LINKS *Linking Dynamic Data*

MAGNET *Machine learning in Information NETWORKS*

SIGMA *Signal, Models and Applications*

Sequel *Sequential Learning*

L'intelligence des données s'intéresse à la manière dont des données sont acquises, manipulées, gérées, enrichies, traitées, ... afin d'en extraire ou de révéler l'information qu'elles contiennent, cela dans le but de prendre une décision ou d'aider à la prise de décision. Le mot « donnée » doit être entendu dans un sens très large ; les données proviennent de sources de différents types (capteurs, bases de données, réseaux sociaux, etc), combinant des informations de différents types (images, textes en langue naturelle, nombres, etc). Dans ce contexte, les données doivent être organisées ; des volumes de données énormes doivent être pris en compte ; les données sont souvent disponibles sous forme de flux dont les caractéristiques peuvent changer au cours du temps ; l'information utile doit être extraite de ce flux de données. Le champ d'applications est extrêmement étendu : si les mondes des affaires et de la finance sont bien connus, les applications abondent dans les sciences (physique, astronomie, biologie, médecine, ...).

Dans CRIStAL, le groupe « intelligence des données » (DatIng) est composé de 4 équipes de recherche : Links, Magnet, Sequel, et Sigma. DatIng combine des compétences fortes en bases de données, apprentissage automatique, fouille de données et traitement de signal. Links s'intéresse aux collections de bases de données liées et aux requêtes liées les pour en extraire des informations. Magnet s'intéresse aux graphes d'objets, incluant notamment des textes en langue naturelle. Sigma s'intéresse aux

► ~ 40 permanents.

DatIng : Data Intelligence Group
Équipe SIGMA
Signal, Models and Applications

ÉQUIPE PUBLICATIONS SÉMINAIRES VIE D'ÉQUIPE LOGIN ESPACE PRIVÉ Recherche

DATING: DATA INTELLIGENCE GROUP (SIGMA)

DatIng : Data Intelligence Group
Équipe SIGMA (Signals, Models and Applications)

Les compétences de l'équipe SIGMA reposent sur des fondements solides en statistiques et en géométrie, ses membres s'intéressent aux objets mathématiques extraits de mesures physiques comme par exemple les images, la lumière polarisée, les sons, les vidéos or encore les capteurs chimiques ou de turbulence. Ces signaux et leurs propriétés sont traités en utilisant le formalisme Bayésien, la théorie des fonctions de croyance, la géométrie de l'information ou la théorie de l'information.

Avec pour dénominateur commun le traitement et la modélisation statistique des signaux, nos activités de recherche s'inscrivent dans les communautés relevant de l'inférence Bayésienne, de la théorie de la décision évidentielle, de l'imagerie polarimétrique ou de la sécurité de l'information.

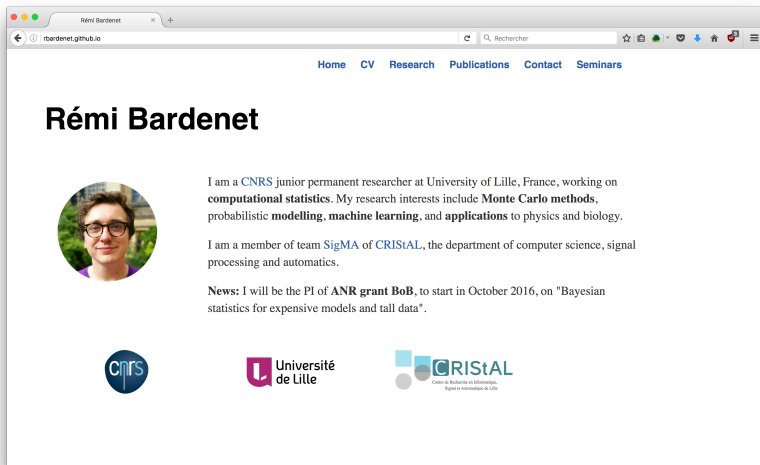
L'équipe SIGMA s'appuie également sur des relations fortes avec le master Decision et Analyse de Données de l'École Centrale de Lille, et l'option de traitement statistique du signal de Telecom-Lille.

Illustrations de nos recherches:

- Inférence Bayésienne et méthodes de Monte-Carlo: suivi d'objets vidéo, modélisation inverse (remontée aux données sources)
- Fusion de données: prise de décision à partir d'un réseau de capteurs
- Sécurité de l'information: détection d'intrusion et de signaux faibles; authentification automatique de documents imprimés
- Imagerie polarimétrique: détection d'artefacts sur des matériaux.

ACTUALITÉS
Le site web est en ligne
La science en marche!

► 13 permanents.



The screenshot shows a web browser window with the URL `rbardenet.github.io`. The page features a navigation menu with links for Home, CV, Research, Publications, Contact, and Seminars. The main content area is titled "Rémi Bardenet" and includes a circular profile picture of Rémi Bardenet. Below the picture, there are three paragraphs of text: a bio mentioning his role as a CNRS junior permanent researcher at the University of Lille, his research interests in computational statistics, Monte Carlo methods, probabilistic modelling, machine learning, and applications to physics and biology; a statement of his membership in the SigMA team at CRIStAL; and a news item about his role as the PI of an ANR grant BoB starting in October 2016. At the bottom of the page, there are three logos: CNRS, Université de Lille, and CRIStAL (Centre de Recherche en Informatique, Signal et Automatique de Lille).

The screenshot shows the Doctissimo Bébé website interface. At the top, there is a navigation bar with links for 'CLUB DOCTISSIMO', 'DEVENIR MEMBRE', 'SE CONNECTER', 'FORUMS', and 'CHAT'. A search bar is present with the text 'Rechercher un article'. Below this is a horizontal menu with categories: SANTÉ, MÉDICAMENTS, GROSSESSE, BÉBÉ, BEAUTÉ, FORME, NUTRITION, RECETTES, FAMILLE, ANIMALIX, PSYCHO, SEXO, VIDÉOS, and TESTS. A secondary menu lists various topics like 'Alimentation de bébé', 'Attaquements', 'Bébé semaine par semaine', 'Biberon', 'Dodo de bébé', 'Diversification', 'Essai de bébé', 'Machettes de bébé', 'Pipi au lit', 'Protonis', 'Protonis', 'Pulvérisateurs BOO', 'Rizome de bébé', 'Somni de bébé', 'Sommeil de bébé', and 'Vaccinations de bébé'. The main content area features an article titled 'Le sommeil de bébé' with a sub-header 'Les besoins de bébé'. To the left, there are sections for 'Dossiers' (including 'CHOISIR UN PRÉNOM' and 'VIDÉOS BÉBÉ'), 'GRANDS DOSSIERS BÉBÉ', and 'BÉBÉ DE 0-3 ANS'. To the right, there is a 'NOS NEWSLETTERS' section with a sign-up form and a 'Gérer mes newsletters' link. At the bottom, a dark footer contains a privacy notice and a 'J'ai compris' button.

Le sommeil de bébé - Docti...
www.doctissimo.fr/html/grossesse/bebe/sommeil-bebe.htm

Rechercher

UN ARTICLE Un médicament

Rechercher un article

SANTÉ MÉDICAMENTS GROSSESSE BÉBÉ BEAUTÉ FORME NUTRITION RECETTES FAMILLE ANIMALIX PSYCHO SEXO VIDÉOS TESTS

Alimentation de bébé
Attaquements
Bébé semaine par semaine
Biberon
Dodo de bébé
Diversification
Essai de bébé
Machettes de bébé
Pipi au lit
Protonis
Protonis
Pulvérisateurs BOO
Rizome de bébé
Somni de bébé
Sommeil de bébé
Vaccinations de bébé

FORUMS BÉBÉ

Accueil / Bébé / Sommeil de bébé

Dossiers

CHOISIR UN PRÉNOM

- Guide des prénoms
- Tous les prénoms

VIDÉOS BÉBÉ

- Premiers pas de bébé
- Bien nourri bébé
- Bébé en pleine forme

GRANDS DOSSIERS BÉBÉ

- Dico bébé
- Baby blog
- Bébé semaine par semaine
- Santé de bébé
- Mode de garde

BÉBÉ DE 0-3 ANS

- Bébé 1 mois

Le sommeil de bébé

Le sommeil de bébé est une préoccupation majeure des jeunes parents. Quel lit choisir ? Dort-il trop ? Pas assez ? Pourquoi pleure-t-il tous les soirs à la même heure ? Quand va-t-il faire ses nuits ? Comment trouver le bon rythme ?... Découvrez tous les secrets du dodo de bébé, pour dormir sur vos deux oreilles !

Les besoins de bébé

Les enfants sont de gros dormeurs. Mais leurs besoins évoluent rapidement en fonction de l'âge. Les parents doivent sans cesse s'adapter à ces changements de rythme. Comment choisir le couchage ? Faut-il le faire dormir dans la chambre des parents ? Quelques informations et conseils pour faciliter le dodo de vos bambins.

NOS NEWSLETTERS

Pour recevoir nos newsletters ou modifier votre inscription, veuillez saisir votre email

Votre email ok

SUVIS HEBDOMADAIRES PERSONNALISÉS

La newsletter grossesse

La newsletter bébé

Gérer mes newsletters

Doctissimo bébé

En poursuivant votre navigation sur ce site, vous acceptez nos conditions générales d'utilisation et notamment que des cookies soient utilisés afin d'améliorer votre expérience d'utilisateur et de vous offrir des contenus personnalisés. Vous êtes par ailleurs informés que nous mettons en oeuvre un système de détection des bloqueurs de publicité sur ce site. [En savoir plus et paramétrer les cookies.](#)

Bayesian inference

- ▶ A biologist decides on
 - ▶ a likelihood $p(\mathbf{x}|\theta)$,
 - ▶ a prior $p(\theta)$,
- ▶ Then he has implicitly decided on
 - ▶ a posterior $p(\theta|\mathbf{x}) = \frac{p(\mathbf{x}|\theta)p(\theta)}{Z}$.
- ▶ Bayesian inference is all about **computing integrals**

$$\int h(\theta)p(\theta|\mathbf{x})d\theta.$$

- ▶ MCMC samples an ergodic Markov chain $(\theta_t)_{t=1,\dots,T}$ with stationary distribution $p(\cdot|\theta)$, so that when $T \rightarrow \infty$,

$$\sqrt{T} \left[\frac{1}{T} \sum_{t=1}^T h(\theta_t) - \int h(\theta)p(\theta|\mathbf{x})d\theta \right] \xrightarrow[T \rightarrow \infty]{d} \mathcal{N}(0, \sigma^2).$$

- ▶ Sampling $(\theta_t)_{t=1,\dots,T}$ requires T likelihood evaluations.

Bayesian inference

- ▶ A biologist decides on
 - ▶ a likelihood $p(\mathbf{x}|\theta)$,
 - ▶ a prior $p(\theta)$,
- ▶ Then he has implicitly decided on
 - ▶ a posterior $p(\theta|\mathbf{x}) = \frac{p(\mathbf{x}|\theta)p(\theta)}{Z}$.
- ▶ Bayesian inference is all about **computing integrals**

$$\int h(\theta)p(\theta|\mathbf{x})d\theta.$$

- ▶ MCMC samples an ergodic Markov chain $(\theta_t)_{t=1,\dots,T}$ with stationary distribution $p(\cdot|\mathbf{x})$, so that when $T \rightarrow \infty$,

$$\sqrt{T} \left[\frac{1}{T} \sum_{t=1}^T h(\theta_t) - \int h(\theta)p(\theta|\mathbf{x})d\theta \right] \xrightarrow[T \rightarrow \infty]{d} \mathcal{N}(0, \sigma^2).$$

- ▶ Sampling $(\theta_t)_{t=1,\dots,T}$ requires T likelihood evaluations.

Bayesian inference

- ▶ A biologist decides on
 - ▶ a likelihood $p(\mathbf{x}|\theta)$,
 - ▶ a prior $p(\theta)$,
- ▶ Then he has implicitly decided on
 - ▶ a posterior $p(\theta|\mathbf{x}) = \frac{p(\mathbf{x}|\theta)p(\theta)}{Z}$.
- ▶ Bayesian inference is all about **computing integrals**

$$\int h(\theta)p(\theta|\mathbf{x})d\theta.$$

- ▶ MCMC samples an ergodic Markov chain $(\theta_t)_{t=1,\dots,T}$ with stationary distribution $p(\cdot|\mathbf{x})$, so that when $T \rightarrow \infty$,

$$\sqrt{T} \left[\frac{1}{T} \sum_{t=1}^T h(\theta_t) - \int h(\theta)p(\theta|\mathbf{x})d\theta \right] \xrightarrow[T \rightarrow \infty]{d} \mathcal{N}(0, \sigma^2).$$

- ▶ Sampling $(\theta_t)_{t=1,\dots,T}$ requires T likelihood evaluations.

Bayesian inference

- ▶ A biologist decides on
 - ▶ a likelihood $p(\mathbf{x}|\theta)$,
 - ▶ a prior $p(\theta)$,
- ▶ Then he has implicitly decided on
 - ▶ a posterior $p(\theta|\mathbf{x}) = \frac{p(\mathbf{x}|\theta)p(\theta)}{Z}$.
- ▶ Bayesian inference is all about **computing integrals**

$$\int h(\theta)p(\theta|\mathbf{x})d\theta.$$

- ▶ MCMC samples an ergodic Markov chain $(\theta_t)_{t=1,\dots,T}$ with stationary distribution $p(\cdot|\theta)$, so that when $T \rightarrow \infty$,

$$\sqrt{T} \left[\frac{1}{T} \sum_{t=1}^T h(\theta_t) - \int h(\theta)p(\theta|\mathbf{x})d\theta \right] \xrightarrow[T \rightarrow \infty]{d} \mathcal{N}(0, \sigma^2).$$

- ▶ Sampling $(\theta_t)_{t=1,\dots,T}$ requires T likelihood evaluations.

Bayesian inference

- ▶ A biologist decides on
 - ▶ a likelihood $p(\mathbf{x}|\theta)$,
 - ▶ a prior $p(\theta)$,
- ▶ Then he has implicitly decided on
 - ▶ a posterior $p(\theta|\mathbf{x}) = \frac{p(\mathbf{x}|\theta)p(\theta)}{Z}$.
- ▶ Bayesian inference is all about **computing integrals**

$$\int h(\theta)p(\theta|\mathbf{x})d\theta.$$

- ▶ MCMC samples an ergodic Markov chain $(\theta_t)_{t=1,\dots,T}$ with stationary distribution $p(\cdot|\theta)$, so that when $T \rightarrow \infty$,

$$\sqrt{T} \left[\frac{1}{T} \sum_{t=1}^T h(\theta_t) - \int h(\theta)p(\theta|\mathbf{x})d\theta \right] \xrightarrow[T \rightarrow \infty]{d} \mathcal{N}(0, \sigma^2).$$

- ▶ Sampling $(\theta_t)_{t=1,\dots,T}$ requires T likelihood evaluations.

Tall data

- ▶ Assume data are independent conditional on θ ,

$$p(\mathbf{x}|\theta) = \prod_{i=1}^n p(x_i|\theta)$$

- ▶ Can you get **the same** central limit theorem while never evaluating all terms in the product?
- ▶ Yes [1], sometimes using $o(n)$ datapoints per iteration! [2]
- ▶ Unanswered yet: What is the equivalent of stochastic gradient for integration?

Tall data

- ▶ Assume data are independent conditional on θ ,

$$p(\mathbf{x}|\theta) = \prod_{i=1}^n p(x_i|\theta)$$

- ▶ Can you get **the same** central limit theorem while never evaluating all terms in the product?
- ▶ Yes [1], sometimes using $o(n)$ datapoints per iteration! [2]
- ▶ Unanswered yet: What is the equivalent of stochastic gradient for integration?

Metropolis-Hastings

```
MH( $p(x|\theta)$ ,  $p(\theta)$ ,  $q(\theta'|\theta)$ ,  $\theta_0$ ,  $N_{\text{iter}}$ ,  $\mathcal{X}$ )  
1   for  $k \leftarrow 1$  to  $N_{\text{iter}}$   
2        $\theta \leftarrow \theta_{k-1}$   
3        $\theta' \sim q(\cdot|\theta)$ ,  $u \sim \mathcal{U}(0,1)$ ,  
4        $\alpha = \frac{\prod_{i=1}^n p(x_i|\theta')p(\theta')}{\prod_{i=1}^n p(x_i|\theta)p(\theta)} \frac{q(\theta|\theta')}{q(\theta'|\theta)}$   
5       if  $u < \alpha$   
6            $\theta_k \leftarrow \theta'$        $\triangleright$  Accept  
7       else  $\theta_k \leftarrow \theta$      $\triangleright$  Reject  
8       return  $(\theta_k)_{k=1,\dots,N_{\text{iter}}}$ 
```

Metropolis-Hastings

```
MH( $p(x|\theta)$ ,  $p(\theta)$ ,  $q(\theta'|\theta)$ ,  $\theta_0$ ,  $N_{\text{iter}}$ ,  $\mathcal{X}$ )
1   for  $k \leftarrow 1$  to  $N_{\text{iter}}$ 
2        $\theta \leftarrow \theta_{k-1}$ 
3        $\theta' \sim q(\cdot|\theta)$ ,  $u \sim \mathcal{U}_{(0,1)}$ ,
4        $\alpha = \frac{\prod_{i=1}^n p(x_i|\theta')p(\theta')}{\prod_{i=1}^n p(x_i|\theta)p(\theta)} \frac{q(\theta|\theta')}{q(\theta'|\theta)}$ 
5       if  $u < \alpha$ 
6            $\theta_k \leftarrow \theta'$        $\triangleright$  Accept
7       else  $\theta_k \leftarrow \theta$      $\triangleright$  Reject
8       return  $(\theta_k)_{k=1,\dots,N_{\text{iter}}}$ 
```

Metropolis-Hastings

```
MH( $p(x|\theta)$ ,  $p(\theta)$ ,  $q(\theta'|\theta)$ ,  $\theta_0$ ,  $N_{\text{iter}}$ ,  $\mathcal{X}$ )  
1   for  $k \leftarrow 1$  to  $N_{\text{iter}}$   
2        $\theta \leftarrow \theta_{k-1}$   
3        $\theta' \sim q(\cdot|\theta)$ ,  $u \sim \mathcal{U}_{(0,1)}$ ,  
4        $\alpha = \frac{\prod_{i=1}^n p(x_i|\theta')p(\theta')}{\prod_{i=1}^n p(x_i|\theta)p(\theta)} \frac{q(\theta|\theta')}{q(\theta'|\theta)}$   
5       if  $u < \alpha$   
6            $\theta_k \leftarrow \theta'$        $\triangleright$  Accept  
7       else  $\theta_k \leftarrow \theta$      $\triangleright$  Reject  
8       return  $(\theta_k)_{k=1, \dots, N_{\text{iter}}}$ 
```

Metropolis-Hastings

```
MH( $p(x|\theta)$ ,  $p(\theta)$ ,  $q(\theta'|\theta)$ ,  $\theta_0$ ,  $N_{\text{iter}}$ ,  $\mathcal{X}$ )  
1   for  $k \leftarrow 1$  to  $N_{\text{iter}}$   
2        $\theta \leftarrow \theta_{k-1}$   
3        $\theta' \sim q(\cdot|\theta)$ ,  $u \sim \mathcal{U}(0,1)$ ,  
4        $\alpha = \frac{\prod_{i=1}^n p(x_i|\theta')p(\theta')}{\prod_{i=1}^n p(x_i|\theta)p(\theta)} \frac{q(\theta|\theta')}{q(\theta'|\theta)}$   
5       if  $u < \alpha$   
6            $\theta_k \leftarrow \theta'$        $\triangleright$  Accept  
7       else  $\theta_k \leftarrow \theta$      $\triangleright$  Reject  
8       return  $(\theta_k)_{k=1, \dots, N_{\text{iter}}}$ 
```

Metropolis-Hastings

```
MH( $p(x|\theta)$ ,  $p(\theta)$ ,  $q(\theta'|\theta)$ ,  $\theta_0$ ,  $N_{\text{iter}}$ ,  $\mathcal{X}$ )  
1   for  $k \leftarrow 1$  to  $N_{\text{iter}}$   
2        $\theta \leftarrow \theta_{k-1}$   
3        $\theta' \sim q(\cdot|\theta)$ ,  $u \sim \mathcal{U}(0,1)$ ,  
4        $\alpha = \frac{\prod_{i=1}^n p(x_i|\theta')p(\theta')}{\prod_{i=1}^n p(x_i|\theta)p(\theta)} \frac{q(\theta|\theta')}{q(\theta'|\theta)}$   
5       if  $u < \alpha$   
6            $\theta_k \leftarrow \theta'$        $\triangleright$  Accept  
7       else  $\theta_k \leftarrow \theta$      $\triangleright$  Reject  
8       return  $(\theta_k)_{k=1,\dots,N_{\text{iter}}}$ 
```


Subsampling approaches

```
MH( $p(x|\theta)$ ,  $p(\theta)$ ,  $q(\theta'|\theta)$ ,  $\theta_0$ ,  $N_{\text{iter}}$ ,  $\mathcal{X}$ )
1   for  $k \leftarrow 1$  to  $N_{\text{iter}}$ 
2        $\theta \leftarrow \theta_{k-1}$ 
3        $\theta' \sim q(\cdot|\theta)$ ,  $u \sim \mathcal{U}_{(0,1)}$ ,
4        $\alpha = \frac{\prod_{i=1}^n p(x_i|\theta')p(\theta')}{\prod_{i=1}^n p(x_i|\theta)p(\theta)} \frac{q(\theta|\theta')}{q(\theta'|\theta)}$ 
5       if  $u < \alpha$ 
6            $\theta_k \leftarrow \theta'$        $\triangleright$  Accept
7       else  $\theta_k \leftarrow \theta$      $\triangleright$  Reject
8   return  $(\theta_k)_{k=1,\dots,N_{\text{iter}}}$ 
```

Subsampling approaches

```
MH( $p(x|\theta)$ ,  $p(\theta)$ ,  $q(\theta'|\theta)$ ,  $\theta_0$ ,  $N_{\text{iter}}$ ,  $\mathcal{X}$ )
1   for  $k \leftarrow 1$  to  $N_{\text{iter}}$ 
2        $\theta \leftarrow \theta_{k-1}$ 
3        $\theta' \sim q(\cdot|\theta)$ ,  $u \sim \mathcal{U}(0,1)$ ,
4        $\psi(u, \theta, \theta') \leftarrow \frac{1}{n} \log \left[ u \frac{p(\theta)q(\theta'|\theta)}{p(\theta')q(\theta|\theta')} \right]$ 
5        $\Lambda_n(\theta, \theta') \leftarrow \frac{1}{n} \sum_{i=1}^n \log \left[ \frac{p(x_i|\theta')}{p(x_i|\theta)} \right]$ 
6       if  $\Lambda_n(\theta, \theta') > \psi(u, \theta, \theta')$ 
7            $\theta_k \leftarrow \theta'$        $\triangleright$  Accept
8       else  $\theta_k \leftarrow \theta$      $\triangleright$  Reject
9   return  $(\theta_k)_{k=1, \dots, N_{\text{iter}}}$ 
```

► Can we use

$$\Lambda_t^*(\theta, \theta') = \frac{1}{t} \sum_{i=1}^t \log \left[\frac{p(x_i^*|\theta')}{p(x_i^*|\theta)} \right] ?$$

Concentration inequalities

- ▶ Let $\delta > 0$, $\theta, \theta' \in \Theta$. We can find $(t, c_t(\delta))$ such that

$$\mathbb{P}(|\Lambda_t^*(\theta, \theta') - \Lambda_n(\theta, \theta')| \leq c_t(\delta)) \geq 1 - \delta.$$

- ▶ For example, sampling without replacement, we prove [3]

$$c_t(\delta) = \dots \times \sqrt{1 - t/n} \frac{\hat{\sigma}_t}{\sqrt{t}} + \dots \times \frac{C_{\theta, \theta'}}{t}.$$

is valid, where $C_{\theta, \theta'} = \max_{1 \leq i \leq n} |\log p(x_i | \theta') - \log p(x_i | \theta)|$.

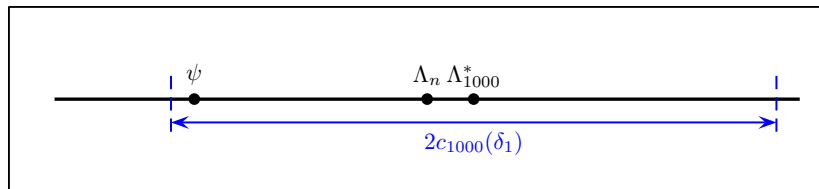
- ▶ **Assume** you can compute $C_{\theta, \theta'}$ in $o(n)$ time.
- ▶ Can we make the right decision with probability $1 - \delta$?

An adaptive choice of t

- ▶ Given $\theta, \theta' \in \Theta$ and $u \in [0, 1]$, an adaptive choice of t can guarantee we know whether

$$\Lambda_n(\theta, \theta) > \psi(u, \theta, \theta')$$

with probability $1 - \delta$.



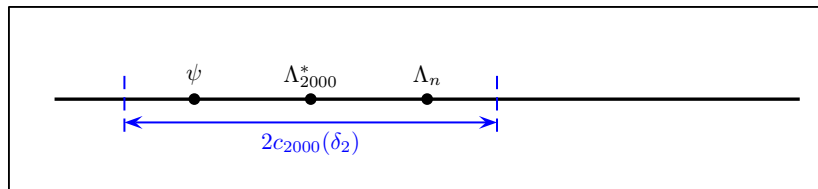
- ▶ Taking (δ_t) such that $\sum_{t \geq 1} \delta_t \leq \delta$ gives the result by a union bound.
- ▶ This Markov kernel inherits the ergodicity of the original MH kernel [1, 4].

An adaptive choice of t

- ▶ Given $\theta, \theta' \in \Theta$ and $u \in [0, 1]$, an adaptive choice of t can guarantee we know whether

$$\Lambda_n(\theta, \theta) > \psi(u, \theta, \theta')$$

with probability $1 - \delta$.



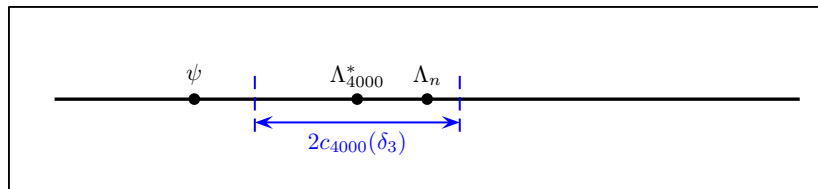
- ▶ Taking (δ_t) such that $\sum_{t \geq 1} \delta_t \leq \delta$ gives the result by a union bound.
- ▶ This Markov kernel inherits the ergodicity of the original MH kernel [1, 4].

An adaptive choice of t

- ▶ Given $\theta, \theta' \in \Theta$ and $u \in [0, 1]$, an adaptive choice of t can guarantee we know whether

$$\Lambda_n(\theta, \theta) > \psi(u, \theta, \theta')$$

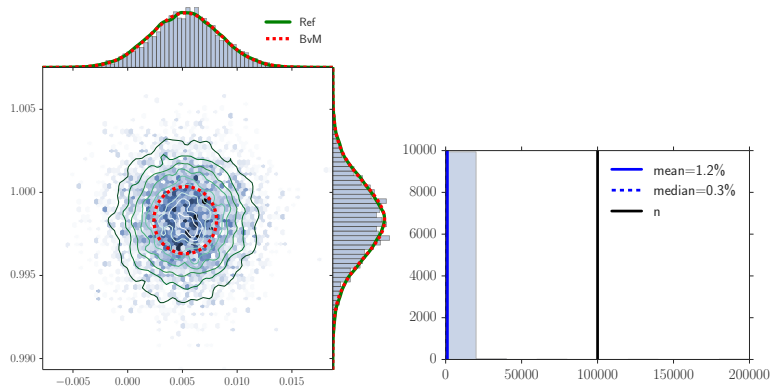
with probability $1 - \delta$.



- ▶ Taking (δ_t) such that $\sum_{t \geq 1} \delta_t \leq \delta$ gives the result by a union bound.
- ▶ This Markov kernel inherits the ergodicity of the original MH kernel [1, 4].

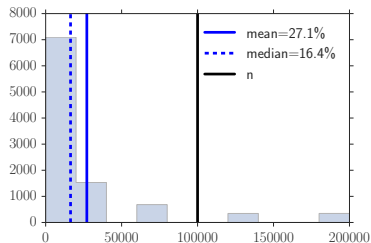
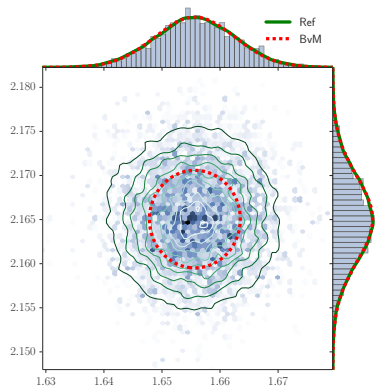
Confidence MH with 2nd order Taylor proxy on a toy example

- ▶ $\mathcal{X} \sim \mathcal{N}(0, 1)$,
- ▶ $p(\cdot|\theta) = \mathcal{N}(\cdot|\mu, \sigma^2)$.



Confidence MH with 2nd order Taylor proxy on a toy example

- ▶ $\mathcal{X} \sim \text{LogNormal}(0, 1)$,
- ▶ $p(\cdot|\theta) = \mathcal{N}(\cdot|\mu, \sigma^2)$.



A saturation phenomenon

- ▶ Toy 2D logistic regression.
- ▶ We can use 2nd order Taylor proxies in this case.

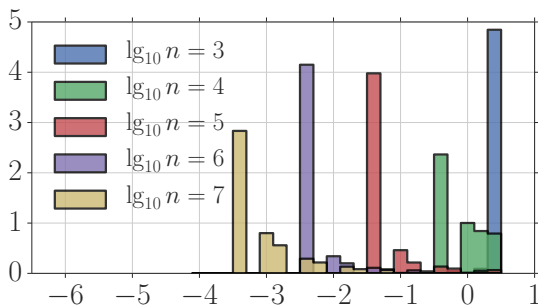


Figure: Histograms of the number of likelihood evaluations

- ▶ We seem to have hit the sample complexity of the problem!

Open issues

- ▶ An explicit relation between δ and the computational budget.
- ▶ Clear recommendations as to when to use Laplace and when to be a proper Bayesian.
- ▶ How do we take into account various data access constraints?
- ▶ Shouldn't we go up one step and start from cost-aware loss functions, then rederive algorithms?

To go further

Our exhaustive (in 2016) review of MCMC for tall data [2]

Pour continuer à communiquer

60G PROBABILITY THEORY AND STOCHASTIC PROCESSES > POINT PROCESSES

Dates - start date: Is greater than or equal to, Title: Is equal to, Apply

La quantification de la répulsion et l'estimation paramétrique pour les processus ponctuels déterminantaux
Christophe Bisio
Aalborg University
30/11/2016 - 14:00
Read more

Géométrie des champs aléatoires de type shot noise: résultats dans le cas Poissonien et perspectives dans le cas déterminantal.
Agnès Desolneux
CNRS-Cachan
25/11/2016 - 13:00
Read more

Systèmes de particules, exposants KPZ, et processus déterminantaux : comment sont-ils reliés ? (suite)
Marielle Simon
INRIA
21/10/2016 - 13:00
Read more

Systèmes de particules, exposants KPZ, et processus déterminantaux : comment sont-ils reliés ?
Marielle Simon
INRIA
21/10/2016 - 11:00
Read more

SHIBBOLETH LOGIN

Shibboleth Login

January

M	T	W	T	F	S	S
				5	6	1
2	3	4			7	8
9	10	11	12	13	14	15
16	17	18	19	20	21	22
23	24	25	26	27	28	29
30	31					

SEARCH SEMINARS BY TAG

- 15A Linear and multilinear algebra; matrix theory > Basic linear algebra
- 35A Partial differential equations > General topics
- 41A Approximations and expansions
- 60B Probability theory and stochastic processes > Probability theory on algebraic and topological structures
- 60G Probability theory and stochastic processes > Point processes
- 62C Statistics > Decision theory
- 62F Statistics > Parametric inference
- 62G Statistics > Nonparametric inference
- 62H Statistics > Multivariate analysis
- 62L Statistics > Sequential methods

UPCOMING SEMINARS

- R. Babbar (Max Planck): TBA
Rohit Babbar
Max-Planck Institute for Intelligent Systems
12/01/2017 - 11:00 to 12:00
- TBA
Benjamin Guedj
Inria MODAL
13/01/2017 - 11:00
- A. Gramfort: TBC
Alexandre Gramfort
Télécom ParisTech, Paris
16/01/2017 - 14:00 to 15:15
- TBA
Olivier Lopez
Université Pierre et Marie Curie
17/01/2017 - 14:00 to 16:00
- TBA
Daniele Calendriello
Inria Sequel
20/01/2017 - 11:00
more

- Les responsables de séminaires peuvent contacter Benjamin Guedj ou moi-même.

References I

- [1] R. Bardenet, A. Doucet, and C. Holmes.
Towards scaling up MCMC: an adaptive subsampling approach.
In Proceedings of the International Conference on Machine Learning (ICML), 2014.
<http://jmlr.org/proceedings/papers/v32/bardenet14-sup.pdf>.
- [2] R. Bardenet, A. Doucet, and C. Holmes.
On Markov chain Monte Carlo methods for tall data.
accepted in Journal of Machine Learning Research (JMLR), 2016.
- [3] R. Bardenet and O.-A. Maillard.
Concentration inequalities for sampling without replacement.
Bernoulli, 2015.
- [4] D. Rudolf and N. Schweizer.
Perturbation theory for Markov chains via Wasserstein distance.
arXiv preprint arXiv:1503.04123, 2015.