# Eléments de processus stochastiques
# Présentation des projets

Maarten Arnst, Vincent Denoël,
Pierre Geurts, Louis Wehenkel

MATH0488-1

Année Académique 2017-2018

Last update : 27 février 2018

# Programme

mardi 6 février 2018, 10.45-12.45 : Intro générale, Chaînes de Markov
mardi 13 février 2018, 10.45-12.45 : Mardi gras
mardi 20 février 2018, 10.45-12.45 : Chaînes de Markov
mardi 27 février 2018, 10.45-12.45 : Proc. 2nd ordre / Intro projets
mardi 6 mars 2018, 10.45-12.45 : Proc 2nd ordre / Simulations
mardi 13 mars 2018, 10 :45-12.45 : projet - séance 1
mardi 20 mars 2018, 10.45-12.45 : projet - séance 2
mardi 27 mars 2018, 10.45-12.45 : projet - séance 3
mardi 3 avril 2018, 10.45-12.45 : congé de printemps
mardi 10 avril 2018, 10.45-12.45 : projet - séance 4 (récup mardi gras)
mardi 17 avril 2018, 10.45-12.45 : projet - séance 5
mardi 24 avril 2018, 10.45-12.45 : projet - séance 6
mardi 1 mai 2018, 10.45-12.45 : projet - séance 7
mardi 8 mai 2018, 10.45-12.45 : projet - séance 8 - Finalisation
mardi 15 mai 2018, 10.45-12.45 : projet - Présentation orale
mardi 22 mai 2018, 10.45-12.45 : Session examen

# Projet

- Projet = problème posé
  - plus ou moins bien balisé,
  - $\nexists$ solution unique
  - $\sim$ dissertation ???
- Projet = aboutissement du cycle de formation "Probabilités, Statistiques, Processus stochastiques"

- Projet à choisir parmi 3 sujets proposés avant le 6 mars 2018 (22h).
- $\exists$ séances de complément théorique
- Groupes de (2-)3 étudiants, importance de travailler ensemble

# Programme des Séances de Projet

Liste des présences **obligatoires**

Projets
(i) S39 du B37 (Inst. Math),
(ii) S42 du B37 (Inst. Math)
(iii) 02 du B37 (Inst. Math)

Répartition des locaux annoncée lorsque les répartitions d'étudiants seront connues.

# Lignes Directrices

**Rapport**

- Un seul rapport par groupe, 15-30 pages, fonte 11pt (figures et bibliographie incluses)
- Biblio - important - citez vos sources
- Code matlab (pas dans le rapport, à fournir en version informatique)
- Remise des rapports le 8 mai 2018 avant 22h00

**Pas de présentation orale**

**Soft skills & Projet**

- Travailler en groupe
- Mener des recherches bibliographiques
- Présenter ses propres résultats de façon concise, orale et écrite
- Utiliser Matlab et découvrir de nouvelles fonctionnalités

# Eléments de processus stochastiques

# Projet : Méthodes de Monte-Carlo par chaines de Markov - application à la cryptanalyse

*Professeurs : Pierre Geurts et Louis Wehenkel*
*Assistante : Laurine Duchesne*

{l.wehenkel, p.geurts, l.duchesne}@uliege.be

`http://www.montefiore.ulg.ac.be/~lduchesne/stocha/`

27/02/2018

# Méthodes de Monte-Carlo par chaînes de Markov (MCMC)

**Problème :**
On veut échantillonner selon une distribution $\pi$.

**Difficulté :**
Souvent, $\pi$ n'est donnée qu'à un facteur de normalisation $\alpha$ près. Si $\alpha$ est impossible à calculer de façon exacte, il est impossible d'échantillonner directement selon la distribution $\pi$.

**Solution :**
Les méthodes MCMC consistent à exploiter une fonction $\gamma(x) = \alpha\pi(x)$ qu'on peut évaluer, afin de simuler une chaine de Markov ergodique dont la *distribution stationnaire* est $\pi$.

# Algorithme de Metropolis-Hastings

Partant d'un état initial $x^{(0)}$, l'algorithme génère une suite d'états $x^{(1)}, x^{(2)}, \ldots$ en appliquant de façon répétitive l'itération suivante:

Given $x^{(t)}$,

1. Generate $y^{(t+1)} \sim q(y|x^{(t)})$
2. Compute $\alpha = \min\left\{1, \frac{\pi(y^{(t+1)})}{\pi(y^{(t)})} \frac{q(y^{(t)}|y^{(t+1)})}{q(y^{(t+1)}|x^{(t)})}\right\}$
3. Set

$$x^{(t+1)} = \begin{cases} y^{(t+1)} & \texttt{with probability } \alpha \\ x^{(t)} & \texttt{with probability } 1-\alpha \end{cases}$$

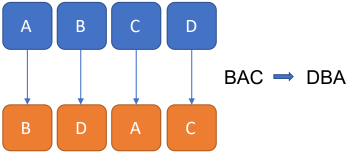
où $q(\cdot|x)$ est une distribution "de proposition" à choisir.

Cela génère une chaine de Markov dont la distribution stationnaire est $\pi(x)$. Le rapport $\pi(x)/\pi(x')$ peut être évalué par $\gamma(x)/\gamma(x')$.

# Le projet

**Objectif général :** Mettre au point un algorithme pour décrypter un message en français encodé par un chiffrement par substitution.

**Chiffrement par substitution :** Chaque symbole de l'alphabet est substitué par un autre, de manière réversible.




Exemple : décoder ''pn,.qeo,gx.qe:éekde',k.oée,fperpreéf:,ggd;rqek,nqe:zbféeo,bnqé-',bnqb.péeqbnekzdbp,n,bpé''

NB: Pour un alphabet de taille $40$, il y a $\approx 2 \times 10^{47}$ chiffrements possibles.

# Le projet

**Idée générale de la solution à mettre en place :**

- Modélisation de l'ensemble des messages $M_n$ possibles par une chaîne de Markov d'ordre 1.
- Sur base d'un texte encodé $D_n = \theta^*(M_n)$, utilisation de l'inférence bayésienne pour trouver le code $\theta^*$ utilisé et le texte $M_n$, en maximisant $P(\theta|D_n)$ étant donnée la modélisation du langage.
- Développement d'un système basé sur un algorithme MCMC pour échantillonner selon $P(\theta|D_n)$ et ainsi déterminer le code $\hat{\theta}$ qui maximise cette probabilité, et puis $\widehat{M}_n = \hat{\theta}^{-1}(D_n)$.
- Analyses des performances de votre système de décryptage sur des séquences de différentes longueurs, avec différents codes...