Markov Chain Monte Carlo et application au rayonnement cosmique

Laurent DEROME

Université Joseph Fourier - LPSC Grenoble

21 mai 2010

Introduction

Quelques remarques préliminaires

- Pas un expert, plus le point de vue d'un utilisateur...
- Un grand nombre de slides de cette présentation sont empruntés à Antje Putze.
- Voir les références à la fin de la présentation.

Parti pris de l'introduction

- MCMC : "simple" générateur de nombres pseudo-aléatoires multidimensionnel.
- A quoi ça peut servir : Méthode MC, Marginalisation d'une PDF.
- A quoi ça sert le plus : Analyse Bayésienne Marginalisation de la PDF $P(\theta|\ \mathrm{data})$

Plan

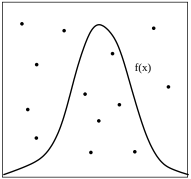
Générateur de VA à 1 dimension

La problématique

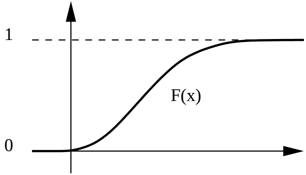
Construire un générateur d'une variable aléatoire dont la PDF est donnée par $f(x)$ à partir d'un générateur uniforme : ingrédient de base des méthodes Monte-Carlo

Deux techniques classiques

- Méthode de la réjection : génération uniforme de points (x, y) et réjection des points $y > f(x)$



- Méthode de la transformation inverse :
  $F(x) = \int_{-\infty}^{x} f(x)dx$
  $x = F^{-1}(u)$ où $u = \mathrm{unif}(0, 1)$



Générateur de VA à m dimensions

La problématique

construire un générateur de m variables aléatoires dont la PDF jointe est donnée par $f(\vec{x})$ où $\vec{x} = (x_1, \ldots, x_m)$à partir d'un générateur uniforme.

On peut toujours utiliser les deux techniques classiques

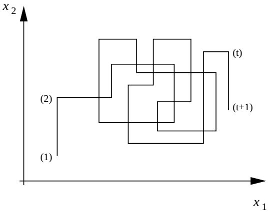
- Méthode de la réjection $\Rightarrow$ Acceptation très faible : problème d'efficacité dés que m est grand ($\sim 1/\alpha^m$)
- Méthode de la transformation il faut calculer (et stoker en mémoire) :
  - $f(x_1)$ (et stocker $F(x_1)$) pour générer x_1
  - $f(x_2|x_1)$ (et stocker $F(x_2|x_1)$ pour chaque x_1) pour générer x_2
  - $f(x_i|\ x_{i-1} \ldots x_1)$ (et stocker $F((x_i|\ x_{i-1} \ldots x_1)$ pour chaque $x_{i-1} \ldots x_1$) pour générer x_i.

  problème de calcul, mémoire dès que m est grand ($\sim \beta^m$)

Générateur de VA à m dimensions : MCMC

le principe

Marche aléatoire dans l'espace des paramètres



Chaine de Markov

Marche aléatoire sans mémoire :

$$P(\vec{x}^{(t+1)}|\vec{x}^{(t)},\ldots,\vec{x}^{(0)}) = P(\vec{x}^{(t+1)}|\vec{x}^{(t)})$$

Méthode MCMC

Chaine de Markov dont les prob. $P(\vec{x}^{(t+1)}|\vec{x}^{(t)})$ sont choisies pour que les $\vec{x}^{(t)}$ soient distribuées selon la PDF $f(\vec{x})$.

Méthode Metropolis

En 2 étapes : proposition et acceptation

On définit

$f(\vec{x})$: fonction cible, c'est la distribution à échantillonner.

$q(\vec{x}|\vec{x}^{(t)})$: loi de proposition, prob. avec laquelle le point $\vec{x}$ est proposé si on est en $\vec{x}^{(t)}$ (loi simple à échantillonner). Dans la méthode de Métropolis, on doit choisir q tel que :

$$q(\vec{x}|\vec{y}) = q(\vec{y}|\vec{x})$$

$a(\vec{x}, \vec{x}^{(t)})$: loi d'acceptation, prob. d'accepter le point proposé $\vec{x}$.

$$a(\vec{x}, \vec{x}^{(t)}) = \begin{cases} 1 & \text{si} \quad f(\vec{x}) > f(\vec{x}^{(t)}) \\ \frac{f(\vec{x})}{f(\vec{x}^{(t)})} & \text{sinon} \end{cases} = \min\left(1, \frac{f(\vec{x})}{f(\vec{x}^{(t)})}\right)$$

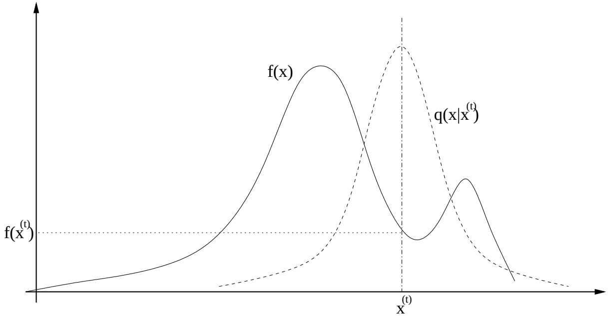
Important

la construction de la chaîne ne dépend pas de la normalisation de $f(\vec{x})$

Méthode Metropolis

En 2 étapes : proposition et acceptation

Exemple à 1D



Méthode Metropolis-Hasting

En 2 étapes : proposition et acceptation

On définit

$f(\vec{x})$: fonction cible, c'est la distribution à échantillonner.

$q(\vec{x}|\vec{x}^{(t)})$: loi de proposition, prob. avec laquelle le point $\vec{x}$ est proposé si on est en $\vec{x}^{(t)}$ (loi simple à échantillonner). Ici on peut choisir q tel que :

$$q(\vec{x}|\vec{y}) \neq q(\vec{y}|\vec{x})$$

$a(\vec{x}, \vec{x}^{(t)})$: loi d'acceptation, prob. d'accepter le point proposé $\vec{x}$.

$$a(\vec{x}, \vec{x}^{(t)}) = \min\left(1, \frac{f(\vec{x})}{f(\vec{x}^{(t)})} \frac{q(\vec{x}^{(t)}|\vec{x})}{q(\vec{x}|\vec{x}^{(t)})}\right),$$

Important

la construction de la chaîne ne dépend pas de la normalisation de $f(\vec{x})$

Convergence de la chaine

La convergence de la marche aléatoire vers la loi de distribution $f(\vec{x})$ est assurée par la condition (detailed balance) :

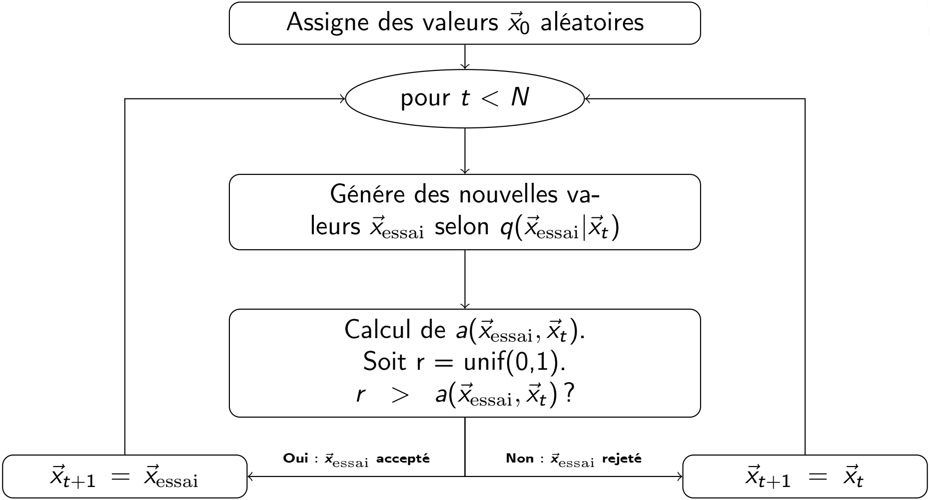
$$P_e(\vec{y}|\vec{x})f(\vec{x}) = P_e(\vec{x}|\vec{y})f(\vec{y})$$

où $P_e(\vec{y}|\vec{x})$ est la prob. de la chaine d'aller en $\vec{y}$ depuis $\vec{x}$. Ici on a donc :

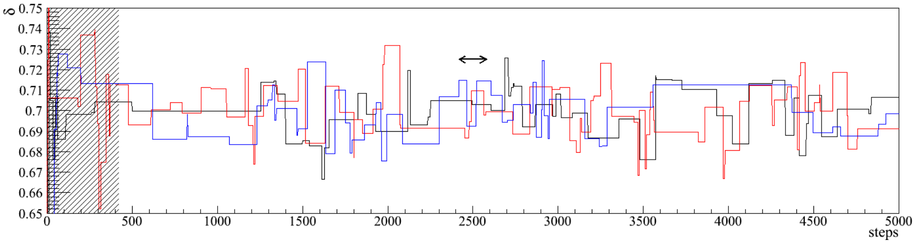
$$P(\vec{y}|\vec{x}) = q(\vec{y}|\vec{x})a(\vec{y},\vec{x})$$

Cette condition est assurée par le choix de la forme de $a(\vec{y}|\vec{x}) = \min\left(1, \frac{f(\vec{y})}{f(\vec{x})}\right)$ (Metropolis) quelque soit le choix de q (symétrique).

Algorithme


Assigne des valeurs $\vec{x}_0$ aléatoires
pour $t < N$
Génére des nouvelles valeurs $\vec{x}_{\rm essai}$ selon $q(\vec{x}_{\rm essai}|\vec{x}_t)$
Calcul de $a(\vec{x}_{\rm essai}, \vec{x}_t)$.
Soit r = unif(0,1).
$r > a(\vec{x}_{\rm essai}, \vec{x}_t)$?
Oui : $\vec{x}_{\rm essai}$ **accepté**
Non : $\vec{x}_{\rm essai}$ **rejeté**
$\vec{x}_{t+1} = \vec{x}_{\rm essai}$
$\vec{x}_{t+1} = \vec{x}_t$

Finalement voila ce qu'un MCMC produit...



- Convergence vers $f(\vec{x})$ mais on doit écarter les premières valeurs prises par la chaîne : Burn-in
- Corrélation entre les valeurs de la chaîne, il faudra garder 1 valeur sur l_c où l_c est la longueur de corrélation pour avoir des échantillons
- Le choix de la loi de proposition est un choix critique :
  - Méthode optimale (n_c petit) : $q(\vec{x}, \vec{x}_t) = f(\vec{x})$.
  - $q(\vec{x}, \vec{x}_t)$ trop étalée $\Rightarrow$ acceptance très faible $\Rightarrow$ l_c grand.
  - $q(\vec{x}, \vec{x}_t)$ trop étroite $\Rightarrow$ corrélation forte $\Rightarrow$ l_c grand.

Pourquoi faire ?

Monte-Carlo

Finalement, à l'aide d'un MCMC on va produire une séquence (indépendante)

$$\vec{x}^1, \vec{x}^2, \ldots$$

distribuée selon la loi cible $f(\vec{x})$.

Marginalisation Estimation de la PDF 1D (ou 2D) :

$$f_p(x_p) = \int_1 \ldots \int_{p-1} \int_{p+1} \ldots \int_m f(\vec{x}) dx_1 \ldots dx_{p-1} dx_{p+1} \ldots dx_m$$

$\Rightarrow$ histogramme des $x_p^1, x_p^2, \ldots$.

Estimation de la valeur de $g(\vec{x})$:

$\Rightarrow$ histogramme des $g(x_p^1), g(x_p^2), \ldots$.

Le cas classique d'utilisation du MCMC : Analyse bayésienne

Estimation des m paramètres $\boldsymbol{\theta} = \left\{\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(m)}\right\}$ d'un modèle théorique à l'aide de données expérimentales

$$\underbrace{P(\boldsymbol{\theta}|\text{données})}_{\text{probabilité } a\ posteriori} \propto \underbrace{P(\text{données}|\boldsymbol{\theta})}_{\text{fonction de vraisemblance}} \cdot \underbrace{P(\boldsymbol{\theta})}_{\text{probabilité } a\ priori}$$

Toute l'information est contenue dans $P(\boldsymbol{\theta}|\text{données})$, PDF à m dimensions des paramètres du modèle. Mais en pratique pour exploiter l'information il faut construire les PDF 1D, 2D : Intégrale multidimensionnelle

$\Longrightarrow$ Méthodes MCMC (fonction cible : $P(\boldsymbol{\theta}|\text{données})$) vont permettre d'évaluer les PDF 1D,2D, …associées aux paramètres compte-tenu des données expérimentales.

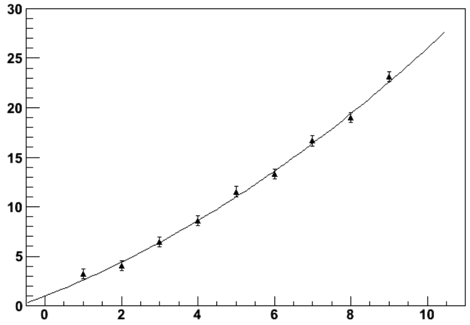
Un exemple simple

Ajustement d'une fonction du second ordre

On dispose de données expérimentales (x_i, y_i) et on a le modèle :

$$y = a + bx + cx^2$$

où $\boldsymbol{\theta} = \{a, b, c\}$ sont les paramètres du modèle. ici on a pris $\boldsymbol{\theta}_{\text{vrai}} = \{a, b, c\}_{\text{vrai}} = \{1, 1.5, .1\}$



Objectif

utiliser les données expérimentales pour reconstruire les PDF associées aux paramètres (a, b, c).

Un exemple simple...

Ajustement d'une fonction du second ordre

Pour cela on écrit :

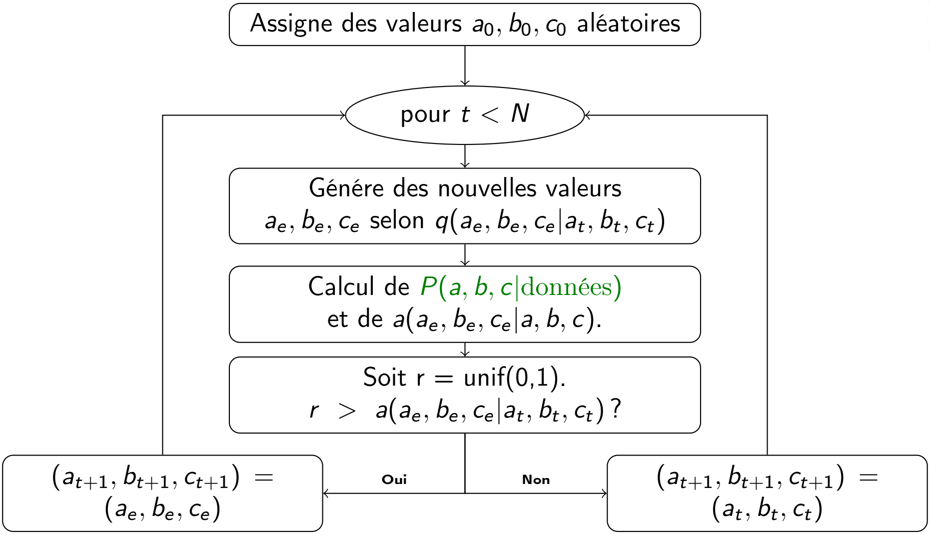
$$P(a,b,c|\text{données}) \propto P(\text{données}|a,b,c) \cdot P(a,b,c)$$

où $P(\text{données}|a,b,c) = \mathcal{L}(a,b,c)$ est la fonction de vraisemblance.
on a :

$$\begin{aligned}\mathcal{L}(a,b,c) &= \Pi_i \left\{ \frac{1}{\sqrt{2\pi}\sigma} \exp -\frac{1}{2}\frac{(y_i - a + bx_i + cx_i^2)^2}{\sigma^2} \right\} \\ &\propto \exp -\frac{1}{2}\chi^2(a,b,c) \quad \left(\chi^2(a,b,c) = \sum_i \frac{(y_i - a + bx_i + cx_i^2)^2}{\sigma^2} \right)\end{aligned}$$

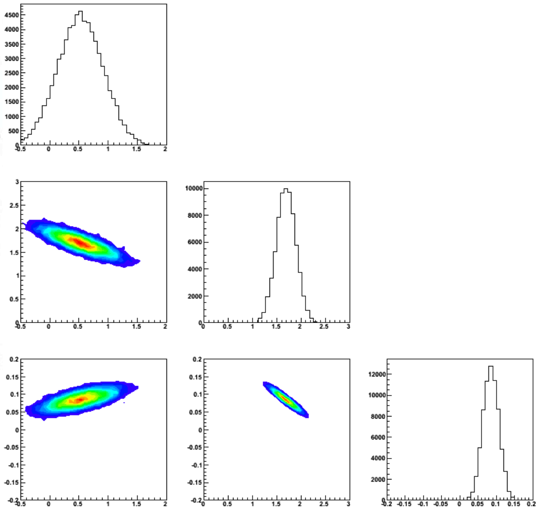
et on choisit $P(a,b,c) = cst$.

Algorithme


Assigne des valeurs a_0, b_0, c_0 aléatoires
pour $t < N$
Génére des nouvelles valeurs
a_e, b_e, c_e selon $q(a_e, b_e, c_e|a_t, b_t, c_t)$
Calcul de $P(a, b, c|\text{données})$
et de $a(a_e, b_e, c_e|a, b, c)$.
Soit r = unif(0,1).
$r \;>\; a(a_e, b_e, c_e|a_t, b_t, c_t)$?
Oui
Non
$(a_{t+1}, b_{t+1}, c_{t+1}) = (a_e, b_e, c_e)$
$(a_{t+1}, b_{t+1}, c_{t+1}) = (a_t, b_t, c_t)$

Un exemple simple

Ajustement d'une fonction du second ordre

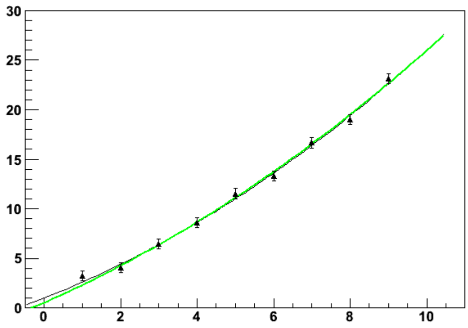


Et on obtient

- PDF 1D des paramètres $\{a, b, c\}$ ($\{a, b, c\}_{\mathrm{vrai}} = \{1, 1.5, .1\}$) : valeur moyenne, erreur, intervalle de confiance
- PDF 2D : corrélations, intervalle de confiance.

Un exemple simple

Ajustement d'une fonction du second ordre



Et on peut aussi construire

- PDF d'observables construites à partir de a, b, c, par exemple $f(x_{\mathrm{new}})$.
- Enveloppe à 68 % de niveau de confiance.
- ...

Reste à comprendre

- Construction de la chaine : comment choisir la fonction de proposition ?
- Contrôle de la convergence de la chaine.
- Analyse de la chaine : Burn-in et longueur de corrélation

Et surtout

Les performances. . . problème pratique clé (en général dominé par le temps de calcul d'un modèle)

$\Rightarrow$ Etude sur un cas physique : propagation du rayonnement cosmique dans la Galaxie

Le rayonnement cosmique

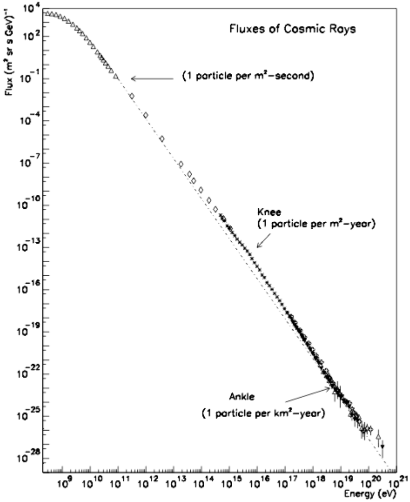
Le spectre énergétique

Le spectre énergétique

- s'étend sur plus de 12 ordres de grandeurs en énergie et 32 ordres de grandeurs en intensité ;
- peut être décrit par une loi de puissance :

$$\frac{\mathrm{d}N(E)}{\mathrm{d}E} \propto E^{-\gamma};$$

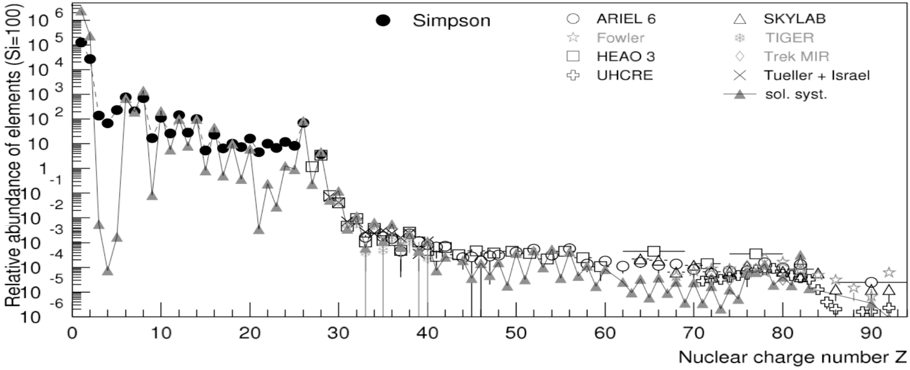
- possède des « anomalies » :
  1. le genou à $\sim 4,5 \times 10^{15}\,\mathrm{eV}$;
  2. la cheville à $\sim 4 \times 10^{18}\,\mathrm{eV}$;
  3. une coupure à $\sim 4 \times 10^{19}\,\mathrm{eV}$.



[Swordy, Space Science Reviews 99 (2001), 85]

Le rayonnement cosmique

Les abondances

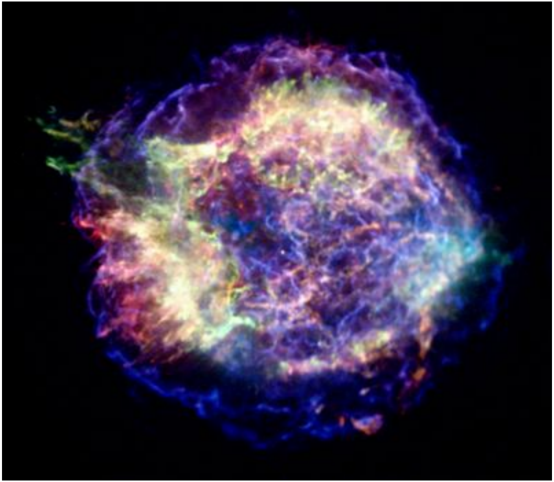


[Hrandel, Advances in Space Research 41 (2008), 442]

87 % hydrogène
12 % hélium
1 % noyaux plus lourds

abondances similaires à celles de notre système solaire, mais
surabondance pour les noyaux $Z = 3 - 5, 20 - 25$

Le périple d'un rayon cosmique galactique

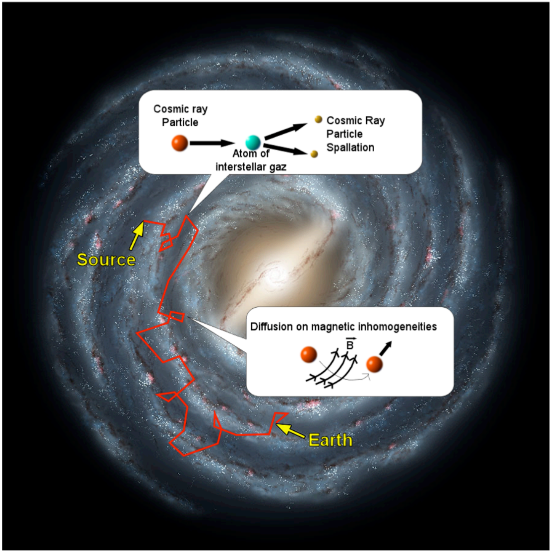


Cassiopée A (Chandra, rayons X), rémanent d'une supernova le plus jeune de la Voie Lactée
[NASA/CXC/MIT/UMass Amherst/M.D.Stage et al.]

Sources - Accélération

étoiles, environnement de supernova ?

Le périple d'un rayon cosmique galactique



[NASA]

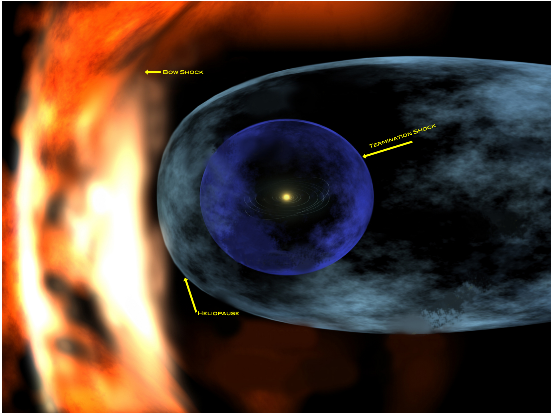
Sources - Accélération

étoiles, environnement de super-nova ?

Propagation dans le milieu inter-stellaire

diffusion sur les inhomogénéités du champ magnétique galactique

Le périple d'un rayon cosmique galactique



L'héliosphére (vue artistique) [NASA]

Sources - Accélération

étoiles, environnement de super-nova ?

Propagation dans le milieu inter-stellaire

diffusion sur les inhomogénéités du champ magnétique galactique

Systéme solaire - Détection

modulation solaire,
coupure géomagnétique

Quelques questions ouvertes

- Quelles et où sont les sources ?
- Existe-t-il des sources exotiques ?
- Comment le rayonnement cosmique est-il accéléré ?
- Comment est-il propagé ?
- Quelle est sa composition à ultra-haute énergie ?
- Quel processus provoque le genou et/ou la cheville ?
- Existe-t-il une limite (GZK ou autre) ?
- ...

Quelques questions ouvertes

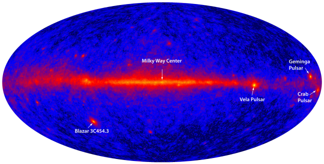
- Quelles et où sont les sources ?
- Existe-t-il des sources exotiques ?
- Comment le rayonnement cosmique est-il accéléré ?
- Comment est-il propagé ?
- Quelle est sa composition à ultra-haute énergie ?
- Quel processus provoque le genou et/ou la cheville ?
- Existe-t-il une limite (GZK ou autre) ?
- ...

Pourquoi étudier la propagation ?

Astrophysique du rayonnement cosmique

L'étude de la propagation du rayonnement cosmique permet de

- déterminer les mécanismes de propagation ;
- contraindre les processus d'accélération ;
- étudier l'émission diffuse des rayons γ.

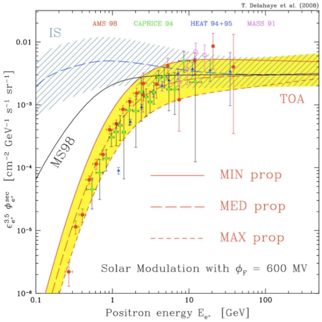


[NASA]

La matiére noire

- annihilation de WIMP χ
  $\chi + \bar{\chi} \rightarrow l\bar{l}, 2\gamma, q\bar{q}, \ldots$
- recherche d'excès dans les spectres de $\bar{p}$, e^+, $\bar{d}$, γ, ...

le « fond » doit être connu



[T. Delahaye *et al.*, A&A 501 (2009), 821]

Questions sur la propagation

4631 (610 MHz)
[Ekers & Sancisi, A&A 54 (1977), 973]

Modèle galactique avec un halo

halo radio dû au rayonnement cosmique autour du disque galactique observé

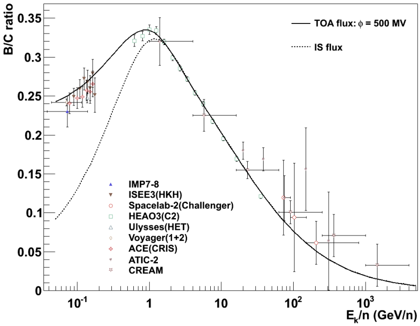
$\Longrightarrow$ halo galactique

Les mécanismes

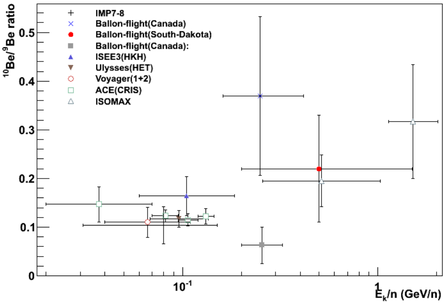
- la diffusion : $K(E)$
  $\Longrightarrow$ le champ magnétique ; Kolmogorov : $K \propto E^{1/3}$?
- la convection : V_c
  $\Longrightarrow$ le vent galactique ;
- la réaccélération : V_a
  $\Longrightarrow$ les ondes magnétohydrodynamiques.

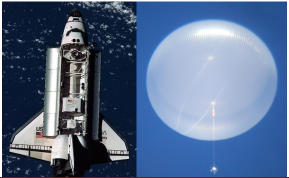
Les observables

Noyaux stables $\Longrightarrow$ matiére traversée



Noyaux radioactifs $\Longrightarrow$ temps de confinement





La détection directe

- mesure des flux élémentaires et isotopiques
- en dehors de l'atmosphère terrestre (satellites, ballons) ;

USINE

Propagation du rayonnement cosmique

Code de propagation USINE développé par D. Maurin

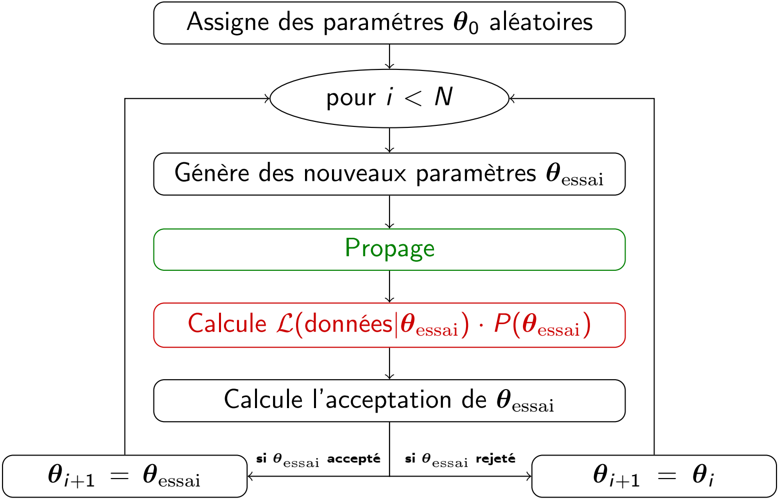
- Résolution semi-analytique de l'équation de diffusion

$$\frac{\partial \psi}{\partial t} = q(\mathbf{r}, p) + \nabla \cdot (D_{xx} \nabla \psi - V\psi) + \frac{\partial}{\partial p} p^2 D_{pp} \frac{\partial}{\partial p} \frac{1}{p^2} \psi - \frac{\partial}{\partial p}\left[\dot{p}\psi - \frac{p}{3}(\nabla \cdot V)\psi\right] - \frac{1}{\tau_f}\psi - \frac{1}{\tau_r}\psi$$

  pour chaque noyau.
- Géométrie à deux zones cylindriques
- Interaction nucléaire - perte d'énergie
- ...
- Produit en sortie tous les flux des RC propagés pour les paramètres données en input.
- Interface avec l'ensemble des mesures du rayonnement cosmique.
- Permet de calculer le χ^2 et $\mathcal{L}$ à partir des observables sélectionnées.

USINE

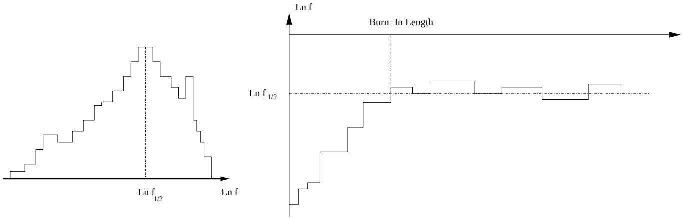
Implémentation du MCMC dans USINE



L'analyse des chaines

Burn-In

On calcule la médiane $f_{1/2}$ de la distribution de la fonction objectif.
La longueur de Burn-in correspond au premier échantillon θ_b pour lequel $f(\theta_b) > f_{1/2}$



L'analyse des chaines

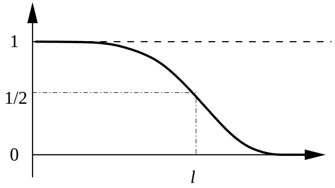
Corrélation

On calcule la longueur d'auto-corrélation pour chaque paramètre $\theta^{(\alpha)}$ $(\alpha = 1, \ldots, m)$:

$$c_j^{(\alpha)} = \frac{\mathrm{cov}\left[\theta_i^{(\alpha)}, \theta_{j+i}^{(\alpha)}\right]}{V\left[\theta_i^{(\alpha)}\right]}.$$

avec la méthode FFT.
La longueur de corrélation $l^{(\alpha)}$: j le plus petit pour lequel $c_j^{(\alpha)} < 1/2$ et

$$l \equiv \max_{\alpha=1,\ldots,m} l^{(\alpha)}.$$



Fonction de proposition

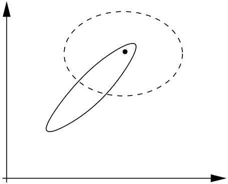
Choix très sensible (surtout pour m grand)

Dans notre cas, on fait 3 MCMC successifs en utilisant le résultat de chaque MCMC pour "construire" une fonction de proposition plus performante.

Etape 1

On a juste une estimation de la précision attendue sur les paramètres, on utilise

$$q\left(\boldsymbol{\theta}_{\text{essai}}|\boldsymbol{\theta}_i\right) \propto \prod_{\alpha=1,\ldots,m} \exp\left(-\frac{1}{2}\frac{\left(\theta^{(\alpha)}_{\text{essai}} - \theta^{(\alpha)}_i\right)^2}{\sigma^2_\alpha}\right).$$



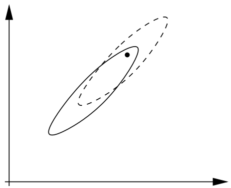
$\Rightarrow$ m gaussiennes indépendantes et centrées en $\boldsymbol{\theta}_i$, $\sigma_\alpha \approx 2.3\sigma_\alpha^{\text{est}}$

Fonction de proposition

Etape 2

On utilise le MCMC précédent pour estimer la matrice de covariance V et on définit

$$q_{COV}\left(\boldsymbol{\theta}_{\text{essai}}|\boldsymbol{\theta}_i\right) \propto \exp\left(-\frac{1}{2}\left(\boldsymbol{\theta}_{\text{essai}} - \boldsymbol{\theta}_i\right)^T V^{-1}\left(\boldsymbol{\theta}_{\text{essai}} - \boldsymbol{\theta}_i\right)\right).$$



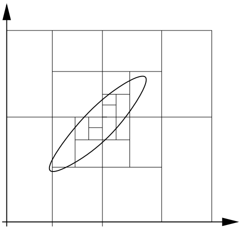
$\Rightarrow$ Permet de proposer des points dans les directions privilégiées

Fonction de proposition

Etape 3

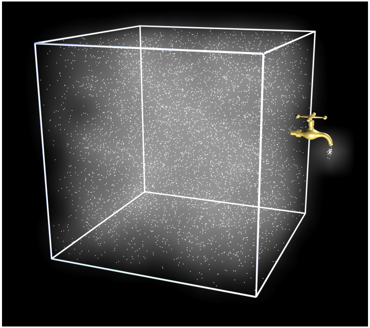
On utilise le MCMC précédent pour construire une fonction de proposition constantes par morceaux sur une partition de l'espace des paramètres construits par division successives de l'espace (Binary Space Partitioning) :

- Une cellule est divisée en deux tant que $N_{\rm cell} > N_{\rm th}$ ($N_{\rm th} \sim 10$).
- la valeur prise par $q_{BSP}\left(\boldsymbol{\theta}_{\rm essai}\right)$ dans chaque cellule est
$$\frac{N_{\rm cell}}{N_{\rm tot}} \frac{1}{V_{\rm cell}}$$
- Fonction $q_{BSP}\left(\boldsymbol{\theta}_{\rm essai}\right)$: approximation de la fonction cible très simple à échantillonner (uniforme par morceau).



Un exemple

Le modéle du Leaky-Box



$$\frac{N_i}{\tau_{\rm esc}} + \bar{n} v \sigma_i N_i = \bar{q}_i + \sum_{j>i} \bar{n} v \sigma_{ij} N_j$$

$$n \quad \Leftrightarrow \quad \bar{n}, \quad q_i \quad \Leftrightarrow \quad \bar{q}_i$$

$$\frac{S}{P} = \frac{\sigma_{PS}}{\bar{m}/\lambda_{\rm esc} + \sigma_S}$$

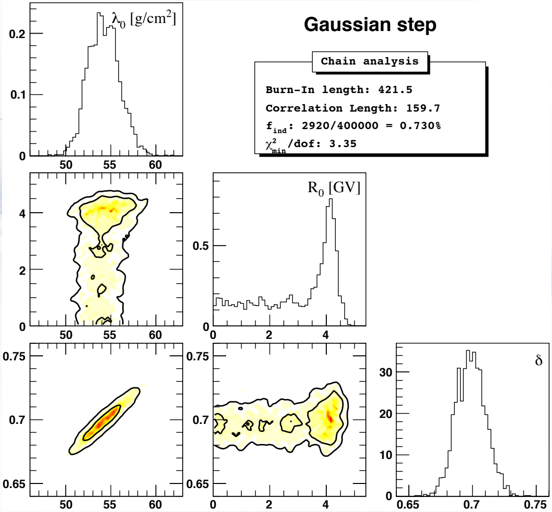
$$\lambda_{\rm esc} = \bar{m}\bar{n} v \tau_{\rm esc}$$

$$\lambda_{\rm esc}(R) = \lambda_0 \beta \begin{cases} {R_0}^{-\delta} & \text{pour } R < R_0, \\ R^{-\delta} & \text{sinon} \end{cases} \qquad \text{avec} \quad R = \frac{pc}{Ze}$$

3 paramètres libres : λ_0 en g cm^{-2}, R_0 en GV, δ

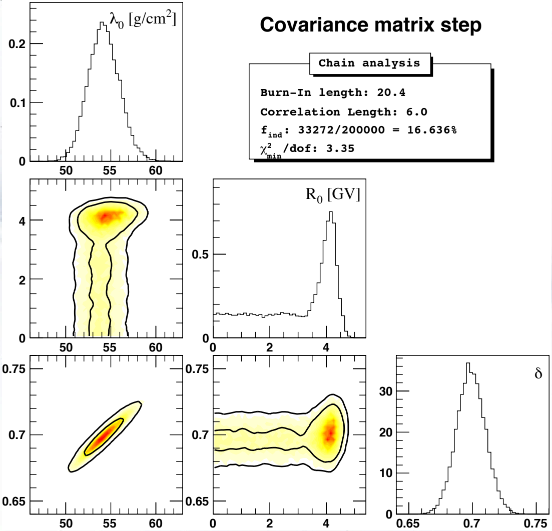
Un exemple

Le modèle du Leaky-Box



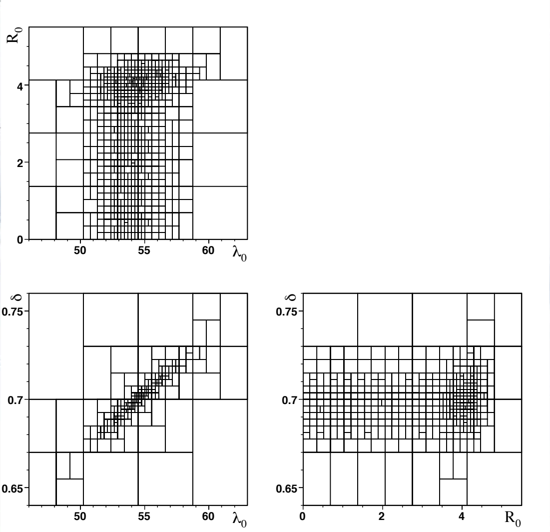
Un exemple

Le modèle du Leaky-Box



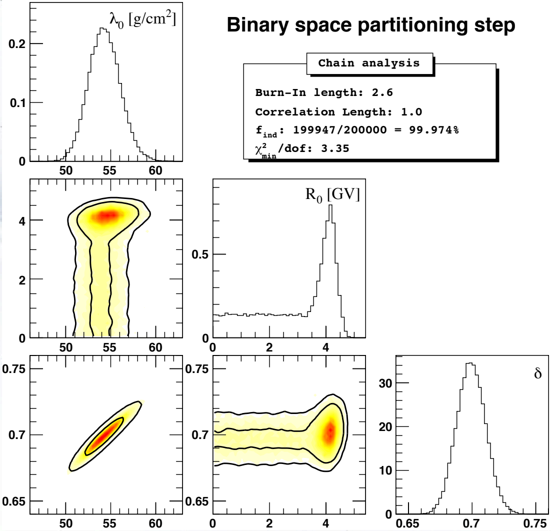
Un exemple

Le modèle du Leaky-Box

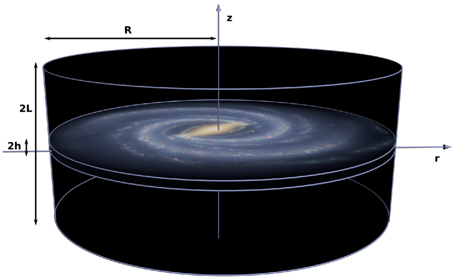


Un exemple

Le modéle du Leaky-Box



Le modèle de diffusion



Galaxie est partagée en deux zones :

1. le disque mince de taille h ;
2. le halo diffusif de taille $L \gg h$.

$$K(R) = K_0 \beta R^\delta$$

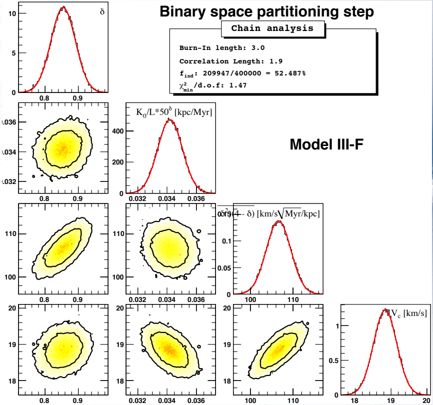
$$n_d = n, \quad n_h = 0$$

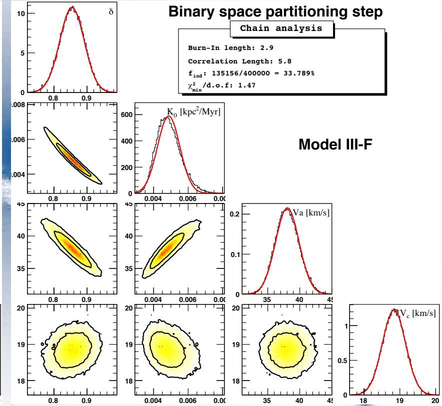
5 paramètres libres : K_0 en kpc^2/Myr, δ, V_c en km/s, L en kpc, V_a en km/s.

Raffinement

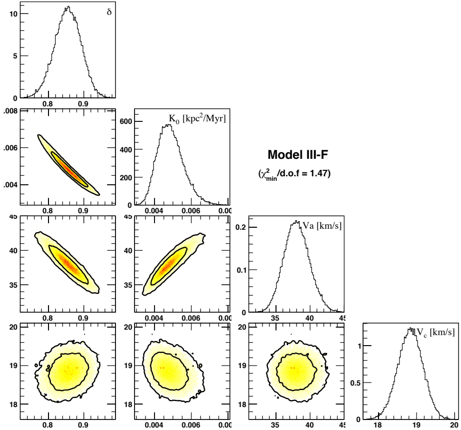
Changement de variable

La performance (longueur de corrélation) dépend beaucoup des corrélations et de la non-gaussianité des PDF : construction de combinaisons des paramètres plus performantes.





Le modèle de diffusion à 1D : les noyaux stables



[Putze *et al.*, A&A 2010]

- Configuration avec convection et réaccélération préférée :

$$L = 4\,\text{kpc fixé}$$
$$V_c = 18,8^{+0,3}_{-0,3}\ \text{km/s}$$
$$\delta = 0,86^{+0,04}_{-0,04}$$
$$K_0 = 0,0046^{+0,0008}_{-0,0006}\ \text{kpc}^2/\text{Myr}$$
$$V_a = 38^{+2}_{-2}\ \text{km/s}$$

- L'indice spectral de Kolmogorov ($\delta = 1/3$) défavorisé par les données utilisées.

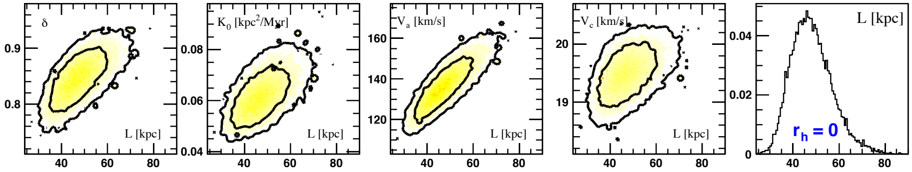
Le modèle de diffusion à 1D : les noyaux radioactifs

Dégénérescence entre K_0 et L pour les rapports primaire sur secondaire stables :

$$\lambda_{\rm esc} = nmvh\frac{L}{K(E)}$$

$\Longrightarrow$ Secondaires radioactifs permettent de lever cette dégénérescence

Résultats avec les données ^{10}Be/^{9}Be

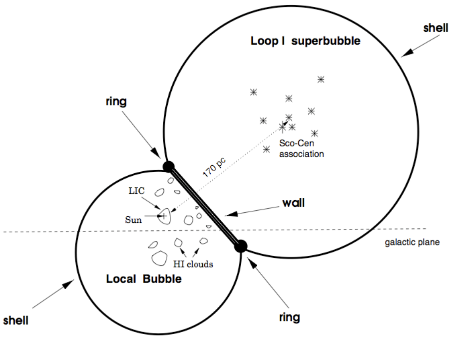


[Putze *et al.*, A&A 2010]

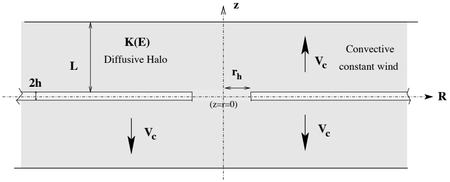
$L = 47^{+12}_{-7}$ kpc

Le modèle de diffusion modifié à 1D

La bulle locale



Modélisation : trou dans le disque



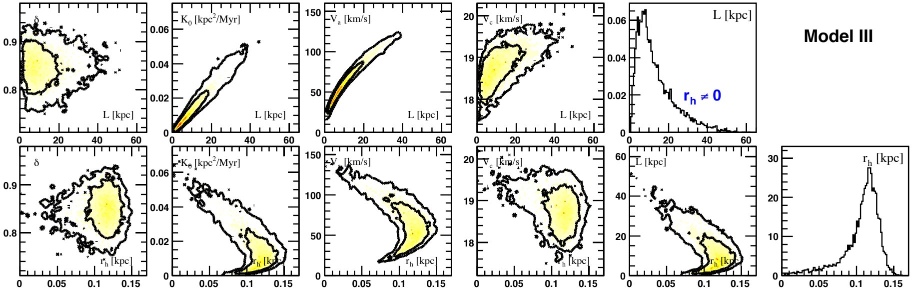
$$\frac{N_{r_h}}{N_{r_h=0}} = \exp\left(\frac{-r_h}{l_{\rm rad}}\right),$$

où $l_{\rm rad}$ est la distance typique parcourue avant la décroissance du radioactif [Donato *et al.*, A&A 381 (2002), 539]

Paramètre supplémentaire : le rayon de la bulle locale r_h en pc

Le modèle de diffusion modifié à 1D : les noyaux radioactifs

Résultats avec les données $^{10}Be/^{9}Be$



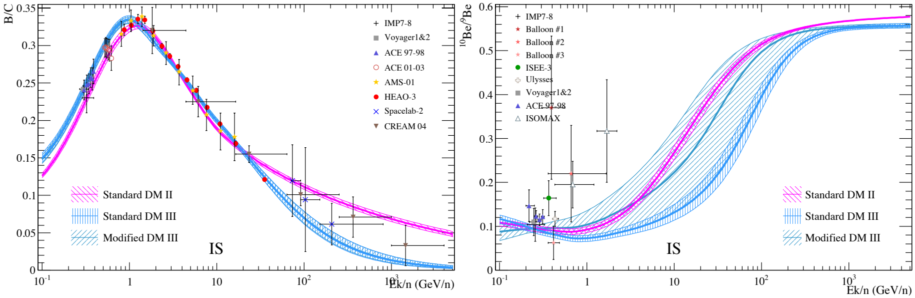
[Putze *et al.*, A&A 2010]

$$L = 8^{+8}_{-7}\,\text{kpc} \qquad r_h = 120^{+20}_{-20}\,\text{pc}$$

Des données plus précises permettront de mieux contraindre L et r_h !

Le modèle de diffusion : Enveloppes

Enveloppes construites avec 68 % des modèles les plus probables



Conclusion : application du MCMC au RC

- Extraction des densités de probabilité *a posteriori* des paramètres de propagation du modèle de diffusion à une dimension :
  - modèle avec réaccélération et convection préféré par les données utilisées ;
  - indice spectral de Kolmogorov défavorisé par les données utilisées ;
- Etude des paramètres de géométrie de la Galaxie :
  - estimation de la taille du halo L et le rayon r_h de la bulle locale ;
  - prise en compte de la bulle locale baisse la taille du halo L ;
  - valeurs trouvées compatibles avec les observations.

Et après

Sélection de Modèle

Problème classique

- Plusieurs modèles concurrents, lequel choisir ?
- Doit-on ajouter un nouveau paramètre libre ?

Etude dans le cadre d'une analyse bayésienne

On a toujours

$$P(\theta|D,H) = \frac{P(D|\theta,H)P(\theta|H)}{P(D|H)},$$

Ici $D \equiv$ data et $H \equiv$ hypothèse/modèle
$P(D|H)$ est l'*évidence* du modèle H, la quantité clé pour la sélection d'un modèle.
On a

$$E = P(D|H) = \int d\theta\, P(D|\theta,H)P(\theta|H)$$

Evidence

- Alors qu'un modèle plus complexe réalisera toujours un meilleur ajustement, l'évidence, étant proportionnelle au volume occupé par le posterior relativement au prior implémente naturellement le "rasoir" d'Occam.
- Il favorise les modèles plus simples et permet de quantifier la tension entre la simplicité d'un modèle et sa capacité à reproduire les données.
- Jeffreys (1961) a fournit un critère pour choisir entre deux modèles :
    - $1 < \Delta \ln E < 2.5$ est substantiel,
    - $2.5 < \Delta \ln E < 5$ est fort,
    - $\Delta \ln E > 5$ est décisif.

Evidence

Calcul de l'évidence

Méthode de calcul

- Nested Sampling : ressemble au MCMC mais échantillonne l'ensemble de l'espace des paramètres de façon plus efficace pour l'intégration.
- Important Sampling

Evidence

Important Sampling

(Ici $P(x)$ est le prior et $\pi(x) = L(x)P(x)$, on a donc :)

$$E = \int L(x)P(x)\,\mathrm{d}x = \int \pi(x)\,\mathrm{d}x,$$

Pour une PDF q dont le support inclus celui de π, on peut écrire :

$$E = \int \pi(x)\,\mathrm{d}x = \int \frac{\pi(x)}{q(x)}\,q(x)\,\mathrm{d}x.$$

La méthode produit une estimation MC de E à partir d'une séquence $x_1, \ldots x_N$ échantillonnée à partir de l' *importance function* q

$$E \approx \frac{1}{N}\sum_{n=1}^{N} w_n; \quad w_n = \frac{\pi(x_n)}{q(x_n)},$$

où w_n sont les poids.

Evidence

Important Sampling

Poids normalisés :

$$\bar{w}_n = \frac{w_n}{\sum_{m=1}^{N} w_m}.$$

Qualité de l'estimation de E dépend du choix de la fonction q. La variance est donnée par

$$\sigma_E^2 = \frac{E^2}{N} d^2(\bar{\pi} \| q),$$

où

$$d^2(\bar{\pi} \| q) = \int \frac{\bar{\pi}^2(x)}{q(x)} \mathrm{d}x - 1.$$

d'autant plus performant que $q(x)$ est proche de $\bar{\pi}(x)$.

Exemple

Implémentation

Dans notre cas : on utilise $q(x) = q_{BSP}(x)$ et la séquence généré $x_1, \ldots x_N$ dans l'étape 3.

Exemple

On va tester différents processus et modèles : un modèle linéaire, un modèle du deuxième ordre et un modèle avec un changement de pente :

- un modèle linéaire

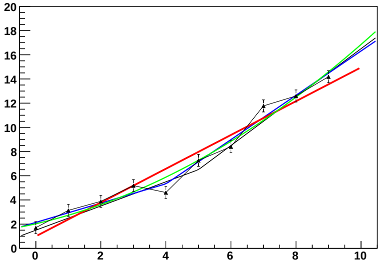
$$f_L = a_1 x + b$$

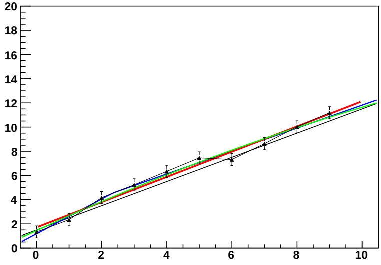
- un modèle du deuxième ordre

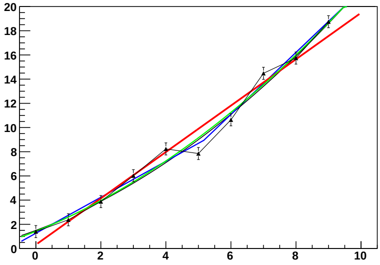
$$f_S = a_1 x + b + c x^2$$

- un modèle avec un changement de pente

$$f_K(x) = \begin{cases} a_1 x + b & \text{pour} \quad x \leq x_K \\ a_2(x - x_K) + a_1 x_K + b & \text{pour} \quad x > x_K \end{cases}$$


20
18
16
14
12
10
8
6
4
2
0
0
2
4
6
8
10


20
18
16
14
12
10
8
6
4
2
0
0
2
4
6
8
10


20
18
16
14
12
10
8
6
4
2
0
0
2
4
6
8
10

Exemple

	Model					
	Kink		Linear		Second	
Process	ln E	χ^2/n.d.f.	ln E	χ^2/n.d.f.	ln E	χ^2/n.d.f.
Kink	-13.5643	10.1734/6	-20.6683	31.855/8	-14.7266	15.9965/7
Linear	-9.17013	4.09954/6	-8.24472	6.65462/8	-9.54192	5.64227/7
Second	-14.8201	12.9957/6	-22.5555	35.4976/8	-15.425	17.4064/7

Evidence vs. χ^2

Ici les avantages de l'évidence sont :

- On dispose d'un critère simple pour sélectionner le modèle
- Peut-être utilisée pour une fonction de vraisemblance quelconque (et pas seulement de la forme $\ln L = -\frac{1}{2}\chi^2$) : petits nombres (poissonienne), barre d'erreurs asymétriques, bin d'énergie très larges, ...
- Prise en compte du Prior, donc plus appropriée à l'approche bayésienne.

Conclusion

- MCMC : générateur de séquence de variable aléatoire.
- Permet d'estimer les PDF marginalisée des paramètres.
- Particulièrement utilisé dans le cadre de l'analyse bayésienne
- Construction de la chaine en théorie très simple
- Analyse de la chaine : Burn-in & corrélation
- Performances : question cruciale.
  - En pratique très sensible au choix de la fonction de proposition.
  - Ici : 3 étapes successives.
  - Existe aussi : algorithme avec auto-adaptation de la fonction de proposition.

Références

MCMC & Application au RC

- R. M. Neal. Probabilistic Inference Using Markov Chain Monte Carlo Methods. Technical Report CRG-TR-93-1, Department of Computer Science, University of Toronto, 1993.
- D. J. C. MacKay. Information Theory, Inference and Learning Algorithms. Cambridge University Press, October 2003.
- "Phénoménologie et détection du rayonnement cosmique nucléaire", A. Putze, Thèse de l'UJF (2009)
- "A Markov Chain Monte Carlo technique to sample transport and source parameters of Galactic cosmic rays - I. Method and results for the Leaky-Box model", Putze A., Derome L., Maurin D., . Perotto L., Taillet R., Astronomy & Astrophysics 497 (2009) 991-1007.
- "A Markov Chain Monte Carlo technique to sample transport and source parameters of Galactic cosmic rays. II. Results for the diffusion model combining B/C and radioactive nuclei" , Putze A., Derome L., Maurin D., accepted for publication in Astronomy & Astrophysics (2010).

Références

Model Selection, Evidence, Nested Sampling, Important Sampling

- "MULTINEST : an efficient and robust Bayesian inference tool for cosmology and particle physics" - F. Feroz, M.P. Hobson and M. Bridges - Mon. Not. R. Astron. Soc. 000, 114 (2008) – astro-ph :0809.3437
- "Bayesian Evidence from Nested Sampling", P. Mukherjee, D. Parkinson and A. R. Liddle – astro-ph :0508461
- "Bayesian model comparison in cosmology with Population Monte Carlo", Martin Kilbinger et al. – astro-ph :0912.1614